From Amazon to Apple: Modeling Online Retail Sales, Purchase Incidence and Visit Behavior

Anastasios Panagiotelis
Michael S Smith, Melbourne Business School
Peter Danaher, Monash University

Available at: http://works.bepress.com/michael_smith/32/
From Amazon to Apple: Modeling Online Retail Sales, Purchase Incidence and Visit Behavior

Anastasios Panagiotelis, Michael S. Smith and Peter J. Danaher

June 13, 2013

Anastasios Panagiotelis is Lecturer in Econometrics at Monash University, Michael S. Smith is Professor of Econometrics in the Melbourne Business School at the University of Melbourne and Peter J. Danaher is Professor of Marketing and Econometrics at Monash University. Correspondence should be directed to Michael Smith at mike.smith@mbs.edu. The authors thank comScore for providing the online panel data, as well as two anonymous referees, an AE, and Editor for comments that helped improve the paper. The work of Michael Smith and Peter Danaher was partially supported by Australian Research Council Grants DP1094289 and FT110100729, and Anastasios Panagiotelis by an Alexander von Humboldt Research Fellowship.
From Amazon to Apple: Modeling Online Retail Sales, Purchase Incidence and Visit Behavior

Abstract

In this study we propose a multivariate stochastic model for website visit duration, page views, purchase incidence and the sale amount for online retailers. The model is constructed by composition from carefully selected distributions, and involves copula components. It allows for the strong nonlinear relationships between the sales and visit variables to be explored in detail, and can be used to construct sales predictions. The model is readily estimated using maximum likelihood, making it an attractive choice in practice given the large sample sizes that are commonplace in online retail studies. We examine a number of top-ranked U.S. online retailers, and find that the visit duration and the number of pages viewed are both related to sales, but in very different ways for different products. Using Bayesian methodology we show how the model can be extended to a finite mixture model to account for consumer heterogeneity via latent household segmentation. The model can also be adjusted to accommodate a more accurate analysis of online retailers like apple.com that sell products at a very limited number of price points. In a validation study across a range of different websites, we find that the purchase incidence and sales amount are both forecast more accurately using our model, when compared to regression, probit regression, a popular data-mining method and a survival model employed previously in an online retail study.

Keywords: Copulas, Marketing Models, Online Purchasing.
1 Introduction

Sales conversion rates for online retailers are notoriously low (Moe and Fader 2004; Venkatesh and Agarwal 2006), with Lin et al (2010) estimating it at 2.3%. This contrasts markedly with offline retailers, who have much higher conversion rates (Moe and Fader 2004). While an online nonsale incurs little monetary cost, there are many businesses that sell exclusively online (e.g., amazon.com, expedia.com and orbitz.com) that would be frustrated by low online conversion rates (Lin et al. 2010). Moreover, even retailers that have both an offline and online presence can realize advantages from increasing their online sales. At the heart of increasing online sales is developing a website that engages the consumer. A simple and often-used measure for engagement with a website is the duration of a visit, sometimes referred to as “stickiness,” which has been linked to online retail profits (Bucklin and Sismeiro 2003; Johnson et al. 2003; Venkatesh and Agarwal 2006). Another related stickiness measure is the number of page views (Danaher et al. 2006), which Manchanda et al. (2006) show is positively related to higher repeat purchase rates by online consumers.

Previous studies have linked duration with purchase incidence (Moe and Fader 2004; Montgomery et al. 2004; Van den Poel and Buckinx 2005), page views with purchase incidence (Manchanda et al. 2006; Van den Poel and Buckinx 2005), both duration and page views to purchase incidence (Lin et al. 2010) and duration to sales amount (Danaher and Smith 2011). However, what is missing from the literature is a simultaneous analysis of purchase incidence, sales amount, visit duration and number of pages viewed. We address this here by developing a quad-variate stochastic model. The marginal distributions of these variables are very different, and the dependence structure highly nonlinear, so that there is no known suitable multivariate distribution to employ. Therefore, we construct one by composition, where the component distributions are flexible and shown to fit well empirically. One component is the bivariate distribution of sales and duration, conditional on page views and purchase incidence, which is captured by a bivariate Gaussian copula model. While copula models are used widely in multivariate modeling (Nelsen 2006; McNeil, Frey and
Embrechts 2005), they have only recently been employed in marketing models; see Danaher and Smith (2011), Glady, Lemmens and Croux (2010) and Park and Gupta (2012) for examples. An advantage of our model is that it can be estimated rapidly using maximum likelihood, even for the very large sample sizes that can occur in online retail studies. In our study we employ data from a panel of approximately 100,000 internet users in the United States, whose internet activity was observed continuously over 2007. We fit our model to datasets from nine of the largest US retail sites, and use this to address a number of research questions.

First, we determine the impact of the website stickiness measures on sales. To do this we derive analytically from our stochastic model the expected sales and probability of purchase, conditional on one or both of visit duration and page views. These conditional expections are difficult to estimate directly using regression style models, both because they are highly nonlinear, and because the four variables are determined simultaneously. We also show how our model can be adjusted to account for online retailers with products at a very limited number of price points, such as apple.com, where the primary sale item is a $0.99 song.

Second, in a validation study we investigate the extent to which our model can be used to forecast both purchase incidence and sales amount, given the website visit variables. For all nine websites examined the forecasts from our model prove to be more accurate using both stickiness measures together, compared to either alone, and a naïve benchmark that uses neither. We also compare the accuracy of forecasts from our model to those from a linear regression and probit model using the stickiness measures as covariates, a popular data-mining method and an adaptation of the survival model developed by Manchanda et al. (2006) for purchase incidence. Throughout, we find our method dominates these alternatives in a statistically significant fashion.

Third, we demonstrate the versatility of our model by using it to investigate empirically the relationships between the visit and sales variables, both within and across different product categories. We examine the relationship for three websites in each of three product
categories: books and digital media, travel services and apparel. Our model reveals that the relationship between sales and website visit duration and page views is complex, nonlinear and occurs in a form that is not readily decomposed in an additive fashion. A number of strong similarities in the relationships for retailers with the same product categories are uncovered. For example, even though amazon.com’s sales are consistently higher than barnesandnoble.com because of higher basket totals, the purchase probability as a function of visit duration and page views is nearly identical at these two websites. However, there can be strong differences across product categories. For example, there is evidence that consumers research products online first, before buying online later, when buying apparel, but not books and digital media. We also find that apple.com is the only website among those in our study where expected sales amounts for purchases decrease when the visit duration is longer than 4 minutes; the relationship is monotonically increasing for all other websites. This reflects the unique goal-directed behavior of customers who visit the apple.com site.

Last, we extend our model to a latent class finite mixture model, where the mixture segments are quad-variate distributions of the type developed here. Estimation is via Bayesian Markov chain Monte Carlo methodology and the number of latent segments is selected using a variant of the Deviance Information Criterion that is robust to label-switching. We find that latent segmentation at the household level improves forecast accuracy for two websites in the validation study, but either does not affect or decreases accuracy for six other sites. However, using the example of oldnavy.com, we show that there is strong evidence of consumer heterogeneity in purchase patterns. We identify three market segments that exhibit features that are consistent with the taxonomy of online consumer behavior discussed by Moe (2003) and Moe and Fader (2004).

2 Modeling Website Visits and Online Sales

In this section we first introduce the data used in the empirical analysis. We then develop and motivate our proposed stochastic model, highlighting some of its advantages over alternative
approaches. We show how to estimate the model, both in the case where sales amounts are considered continuously distributed, as well as for the case where retailers, such as apple.com, offer products mostly at only a few price points. From the model we derive the expected sales amount and the probability of a purchase, conditional on pages viewed and duration of a visit. Last, we use these to derive predictions in a validation study, which we later show dominate forecasts from a number of alternative methodologies.

2.1 ComScore Data

The data used in this study were collected during 2007 by comScore and made available by subscription via the Wharton Research Data Service (WRDS). The database comprises a subsample of approximately 100,000 members of a panel of over two million internet users from across the United States. Members of the panel have software installed on individual machines in a household which passively captures all website visitation and online transaction activity. The smaller subsample is formed in a proprietary manner based on demographic variables so that it is representative of the wider population of US internet users. The domain names of websites visited are recorded, along with the total number of page views \((P)\) at each domain visited and the total duration \((D)\) of the visit to a particular website. An indicator denotes whether each visit results in a purchase \((B = 1)\), or not \((B = 0)\). If one or more items are purchased during the visit, the total sale amount \((S)\) for the basket is also recorded. The data are collected at machine level and present an opportunity to investigate how the relationship between online transactions and visit behavior varies within and across online retailers.

Table 1 gives the top twenty online retail websites during 2007, ranked both by total sale amounts and by the total number of purchases. Unsurprisingly, amazon.com is the top-ranked website for total sales, yet is ranked second behind apple.com for the total number of purchases. However, as we see later, a substantial number of purchases at apple.com correspond to purchases of a single song at the relatively low amount of $0.99, and so the website is only ranked twentieth by total sales. The table reveals that online retail activity
during 2007 was dominated by sales in apparel, print and digital media, travel services, shipping services, photo processing, computing and electronic equipment, homeware and health and beauty products. We examine the first three of these product categories in detail in Section 3, and select the three major websites in each category. These are listed in Table 4, along with the sample sizes for each site, which are large; for example, there are 407,805 visits to amazon.com, and 268,437 to apple.com. However, this represents less than 5% of the activity at these sites of the full comScore panel, and less than 0.1% of activity at these sites by all US internet users.

2.2 Stochastic Model

We model the distribution of the two website visit variables, duration time, $D > 0$, and number of page views, $P \in \{1, 2, 3, \ldots\}$, joint with the purchase indicator, $B \in \{0, 1\}$, and the sale amount $S \geq 0$. At first glance it might appear that each margin can be modeled separately, with dependence captured by a four-dimensional copula function. Copula modeling is a popular method for constructing multivariate distributions in statistical (Nelsen 2006), econometric (Trivedi and Zimmer 2005) and financial (McNeil, Frey and Embrechts 2005) analysis, although they have only been employed recently in the marketing literature (Danaher and Smith 2011; Park and Gupta 2012). They permit the combination of univariate marginal distributions that need not come from the same distributional family, yet still allow for dependence among variables. However, in this situation direct use of a copula model has limitations. First, the sale amount is zero when no purchase is made in a visit, so that $S$ is degenerate at 0 when $B = 0$, and this cannot be accounted for using existing four-dimensional parametric copula functions. Second, the popular Gaussian and t copulas (McNeil, Frey and Embrechts 2005; p.191) only have 6 and 7 parameters, respectively, while most Archimedean copulas only have a single dependence parameter. This proves insufficient to capture the dependence structure between these four variables, which we show is more nuanced in our empirical work.
Instead, we construct the joint distribution via composition as

\[ F(S, B, D, P) = F_1(S, D | B, P) F_2(B | P) F_3(P), \]  

(2.1)

from the component distributions \( F_1, F_2 \) and \( F_3 \). This has a number of advantages. First, with an appropriate choice of component distributions, nonlinearities and other complexities in the dependence between the sales and website visit variables can be uncovered. We derive the purchase probability and expected sales amount, conditional on the visit variables, to provide insight into these relationships. Second, the degeneracy of \( S \) at 0 is easily accounted for by modeling \( F_1 \) differently when \( B = 1 \) or \( B = 0 \). Third, we identify parametric distributions for each component in equation (2.1) that fit our website data well, distributions that have been used to model similar variables previously. Furthermore, some of the components can be modeled semiparametrically, a property that will be exploited to handle retailers that make most sales at a small number of discrete price points. Fourth, estimation of the parameters of the distribution in equation (2.1) is straightforward using maximum likelihood. Last, because equation (2.1) is a fully-specified joint distribution, it can be readily extended to account for market segmentation, as we show in Section 4.

In building our model we first select an appropriate distribution for the total number of page views per visit, denoted \( F_3 \). A popular model for page views is the negative binomial distribution (NBD), which has been used previously by Danaher (2007) and Huang and Lin (2006). The distribution can be derived as a Gamma-Poisson mixture and is robust to observation-level heterogeneity. In our database, each observation corresponds to a visit to a retailer’s website, where at least one page is viewed, so \( P \geq 1 \). To account for this we truncate the NBD by removing the zero case from the probability mass function.

The remaining two distributions \( F_1 \) and \( F_2 \) in the decomposition of equation (2.1) are both conditional on the number of page views. We account for this by making the parameters of the two distributions functions of \( P \). In particular, we partition \( P \) into contiguous intervals
that cover the range of $P$, and allow the parameters of $F_1$ and $F_2$ to differ in each partition, so they are step functions with respect to $P$. We select the partition cut points for each website so that there are approximately an equal number of visits that result in a purchase within each partition. We select the number of partitions using 5-fold cross-validation as discussed in Section 2.3. In our empirical work we show that this greatly enhances the quality of fit, as well as substantially improving prediction accuracy. Conditional on the page view partition, we model the purchase indicator, $F_2(B|P \in \tilde{P}_k)$, with a Bernoulli distribution.

The bivariate distribution of sale amount and duration of visit, denoted $F_1$, differs depending on whether or not a purchase occurs during the visit. When there is no purchase ($B = 0$), the bivariate distribution is degenerate at $S = 0$, so that $F_1(S = 0, D|B = 0, P \in \tilde{P}_k) = F_{1D}(D|B = 0, P \in \tilde{P}_k)$. The distribution $F_{1D}(D|B = 0, P \in \tilde{P}_k)$ is univariate and relates only to the website duration, and is well-modeled as an Inverse Gaussian. This distribution was identified as optimal using AIC and BIC from a list of alternative distributions that also included the Gamma, Weibull, Log-Logistic, and Log-Normal. It has also been used previously to model duration over heterogeneous populations (Hougaard 1984; Johnson, Kotz and Balakrishnan 1994, p.291), which is precisely the situation here.

However, when a purchase does occur, so that $S > 0$, then $F_1$ is a bivariate distribution which we model using a copula model, as now detailed. We label the two univariate distributions as $F_{1S}(S|B = 1, P \in \tilde{P}_k)$ and $F_{1D}(D|B = 1, P \in \tilde{P}_k)$, which makes explicit the conditioning on purchase incidence and page views. We “couple” these together using a bivariate copula function $C$ with dependence parameter $\theta_k$ that varies over partition, so that

$$F_1(S, D|B = 1, P \in \tilde{P}_k) = C(F_{1S}(S|B = 1, P \in \tilde{P}_k), F_{1D}(D|B = 1, P \in \tilde{P}_k); \theta_k). \quad (2.2)$$

We employ the bivariate Gaussian copula, $C(u, v; \theta) = \Phi_2(\Phi_1^{-1}(u), \Phi_1^{-1}(v); \theta)$, where $\Phi_1^{-1}$ is an inverse standard normal distribution function and $\Phi_2$ is the distribution function of
a bivariate normal distribution with zero mean, unit marginal variances and correlation \(-1 < \theta < 1\). The copula only accounts for any dependence between the two variables; with positive dependence when \(\theta > 0\), negative dependence when \(\theta < 0\) and independence when \(\theta = 0\); see Song (2000) for a discussion of the Gaussian copula. Other bivariate copulas \(C\) can also be employed easily here, with comprehensive lists given by Nelsen (2006), McNeil et al. (2005) and Trivedi and Zimmer (2005). The Gaussian copula has symmetric tail dependence (McNeil et al. 2005, pp. 208-222), which proves an unrealistic assumption in risk management applications. To investigate if this is also unrealistic here, we considered a copula constructed as a linear combination of a Clayton, Gumbel and Gaussian copulas. The Clayton copula allows for lower tail dependence, and the Gumbel for upper tail dependence. When estimated using maximum likelihood for the data from amazon.com, we found minor or no meaningful tail dependence for any page view partition, and therefore retain the Gaussian copula for its robustness and simplicity.

For the univariate distribution of visit duration we again employ an Inverse Gaussian, but for sales amount we employ a Log-Logistic distribution. These were identified as optimal, using AIC and BIC, from alternatives that included the Gamma, Inverse Gaussian, Weibull, Log-Logistic and Log-Normal distributions. We note that the Log-Logistic has been used to model sales amount previously by Oyer (2000).

For apple.com, the 77,398 recorded sales occur at only 151 non-zero price points, with 87.33% of all purchases being for exactly $0.99, which corresponds to the sale of a single song from the iTunes store. The next most popular price point, representing 4.76% of total purchases is $9.99, and corresponds to the purchase of an album. Clearly, the sales amount \(S\) does not follow a Log-Logistic distribution, or any other well known parametric distribution. For this retailer, we therefore employ the empirical distribution function (EDF) for \(S\) in each page view partition, giving an estimated distribution function \(\hat{F}_{1S}(s|B = 1, P \in \tilde{P}_k)\) for each \(k\). This is a nonparametric estimator for each ordinal-valued distribution, with values at all the unique observed price points. The ability to combine parametric copula functions
with one or more nonparametric marginal distributions is widely considered a strength of the copula approach to constructing bivariate distributions (Shih and Louis 1995).

Table 2 lists the component distributions in the model, including their probability density or mass functions and unknown parameters. For the websites other than for apple.com, there are 8 parameters for each page view partition and 2 for the modified NBD for the number of page views itself, resulting in \(8K + 2\) parameters in total.

### 2.3 Estimation

When parametric distributions are used for the components in equation (2.1), each component can be estimated using maximum likelihood, and the resulting point estimates also maximize the joint likelihood. The parameters of the distributions of \(F_1\) and \(F_2\) are estimated separately for each partition, and \(F_1\) also for the two values of \(B\). When \(B = 1\), estimation of the bivariate copula model in equation (2.2) is also straightforward using maximum likelihood, as outlined in Cherubini et al. (2004; pp.154-156). The ease with which copula models can be estimated is one reason for their popularity.

In the case of discrete pricing, estimation of the bivariate distribution \(F_1\) differs because \(F_{1S}(S|B = 1, P \in \tilde{P}_k)\) is modeled as a nonparametric discrete distribution using the EDF. The parameters of \(F_{1D}(D|B = 1, P \in \tilde{P}_k)\) are estimated using maximum likelihood, as is each copula parameter \(\theta_k\), but conditional on the two estimated margins as follows. For each page view partition, given the EDF for \(S\) and estimated Inverse Gaussian distribution for duration \(D\), the likelihood can be calculated as follows. Let \((s_i, d_i)\) be the \(i\)th observation of the pair \((S, D)\), \(u_{D,i} = \hat{F}_{1D}(d_i|B = 1, P \in \tilde{P}_k)\), \(b_i = \hat{F}_{1S}(s_i|B = 1, P \in \tilde{P}_k)\) and \(a_i = \hat{F}_{1S}(s_i^-|B = 1, P \in \tilde{P}_k)\) be the left hand limit of the step function \(\hat{F}_{1S}\) at \(s_i\). Then, the mixed density of \((S, D)\) is obtained by differencing with respect to the discrete-valued \(S\), and differentiation with respect to \(D\). The conditional likelihood for partition \(k\) is

\[
f(\theta_k; \text{data}) = \prod_i \left( \tilde{C}(b_i|u_{D,i}; \theta_k) - \tilde{C}(a_i|u_{D,i}; \theta_k) \right) f_{1D}(d_i|B = 1, P \in \tilde{P}_k),
\]
where $\tilde{C}(u_1|u_2; \theta) = \frac{\partial}{\partial u_2} C(u_1, u_2; \theta)$, $f_{1D}$ is an Inverse Gaussian density, and the product is taken only over observations $i$ where sales are made ($B = 1$) and with page views in partition $k$. This likelihood is easily maximized with respect to $-1 < \theta_k < 1$ by grid search. This approach to estimating a bivariate copula model is widely called ‘inference for margins’, which Joe (2005) shows provides consistent estimates with only a very minor reduction in efficiency compared to full maximum likelihood when the marginal distributions are known.

Last, we select the number of partitions $K$ for each website using 5-fold cross-validation (CV). For each fold $j$, the metric is the score $CV_j = \sum_{i \in \text{Fold}_j} \log(\hat{f}(S_i, B_i|P_i, D_i))$, where the summation is over the observations in fold $j$, and the mixed density is from the model fitted using the other four folds; the overall score is $CV = \sum_{j=1}^5 CV_j$. The mixed density is computed as $f(S, B|P, D) = f(S|B, P, D)f(B|P, D)$, where $f(S|B, P, D)$ is derived from the copula model for $F_1$, and $f(B|P, D)$ is computed as in equation (2.3) below. The last column of Table 4 gives the optimal number of partitions in the range 1 to 30 identified by this approach for all nine websites examined in our study.

### 2.4 Putting the Model to Use

Our primary aim is to understand the impact of the visit variables on both sales and purchase incidence. A simple, yet useful, summary measure is the overall (or “marginal”) level of pairwise dependence between all six pairs of variables. Spearman’s rho is an appropriate measure of dependence even when the margins are far from Gaussian, and can be computed for the distribution at equation (2.1) by first simulating many iterates from this distribution, and then computing the correlation coefficient of the ranked iterates. To simulate an iterate, first simulate $P \sim F_3$, then $B \sim F_2$ and then $(S, D) \sim F_1$. For the latter, if $B = 0$ then $S = 0$ and only $D$ needs generating; while if $B = 1$ then the pair are generated from a bivariate Gaussian copula as outlined in Cherubini et al. (2004, p.181). The pairwise dependence between $S$ and $D$, conditional on a purchase made ($B = 1$) and page views $P$, can also be computed for the bivariate Gaussian copula, with Spearman’s rho $\rho^C = (6/\pi)\arcsin(\theta_k/2)$. We label this with a superscript “C” to distinguish it from the marginal Spearman’s rho.
The impact of page views and duration can be assessed by evaluating the expected sales amount and probability of purchase, conditional on the visit variables as follows. The probability of a purchase is

\[
\Pr(B = 1|D, P) = \frac{f_{1D}(D|B = 1, P)\Pr(B = 1|P)}{f_{1D}(D|B = 1, P)\Pr(B = 1|P) + f_{1D}(D|B = 0, P)\Pr(B = 0|P)}.
\]

(2.3)

Here, \( f_{1D}(D|B = 1, P) \) and \( f_{1D}(D|B = 0, P) \) are the two Inverse Gaussian densities computed at point \( D \), and \( \Pr(B|P) \) is the Bernoulli purchase probability.

The expected sales \( E(S|D, P) = \int s f(s|D, P)ds \) is obtained via univariate numerical integration, except in the case of apple.com where the discrete domain of sales is summed over instead. The density (or mass function for apple.com) \( f(s|D, P) \) can be derived analytically from the components as follows. First, note that

\[
f(s|D, P) = f(s|B = 0, D, P)\Pr(B = 0|D, P) + f(s|B = 1, D, P)\Pr(B = 1|D, P),
\]

where \( f(s|B = 0, D, P) \) is a point mass of 1 at \( s = 0 \) and \( \Pr(B = 1|D, P) \) is given at equation (2.3). For continuously distributed sales amounts,

\[
f(s|B = 1, D, P) = \frac{f_1(s, D|B = 1, P)}{f_{1D}(D|B = 1, P)} = c(F_{1S}(s|B = 1, P), F_{1D}(D|B = 1, P); \theta) f_{1S}(s|B = 1, P),
\]

where \( f_1, f_{1D} \) and \( f_{1S} \) are the density functions of distributions \( F_1, F_{1D} \) and \( F_{1S} \), and \( c(u, v; \theta) = \frac{\partial^2 C}{\partial u \partial v} \) is the copula density. The copula density for a bivariate Gaussian copula is given in Table 2. For sales amounts with a discrete distribution

\[
f(s|B = 1, D, P) = \left[ \tilde{C}(b|F_{1D}(D|B = 1, P); \theta) - \tilde{C}(a|F_{1D}(D|B = 1, P); \theta) \right],
\]

(2.4)

where \( b = F_{1S}(s|B = 1, P) \) and \( a = F_{1S}(s^-|B = 1, P) \) is the left hand limit of \( F_{1S} \) at \( s \).
An advantage of computing the conditional expectation $E(S|D,P)$ and probability $Pr(B = 1|D,P)$ from equation (2.1), rather than modeling them explicitly, is that it captures the contemporaneous and nonlinear dependence between all four variables. We use the conditional expectation and purchase probability derived above in our empirical work in Section 3. We also use them to make predictions in a validation study to further motivate our model, as discussed below.

2.5 Model Validation

We demonstrate that our model improves prediction of sales, conditional on both visit variables, compared to some alternative approaches. These include a naïve forecast as a benchmark, linear modeling, a popular data-mining method and a survival model from the marketing literature. We also show that forecasts constructed using both visit variables are more accurate than those that use only one.

We apply these approaches to the nine websites examined in detail in Section 3. The holdout sample size $n_f = 5,000$ for seven sites, and $n_f = 10,000$ for the two sites with the largest number of observations, amazon.com and apple.com. The holdout sample is stratified with respect to sales amount $S$, so that it is more representative than a simple random sample, although the conclusions are the same in either case. We fit our model to the data for each website and compute the probability of a purchase $\hat{b}_i$, and expected sales $\hat{s}_i$, for each observation $i$ in the holdout sample using the expressions in Section 2.4. These are used to predict purchase incidence and sales amount. We label our method ‘SM’, and compare it to alternative methods which construct forecasts as follows:

- **Naïve (N)**: The historical purchase incidence ($\hat{b}_i = \bar{B}$), and the average sales amount $\hat{s}_i = \bar{S}$.

- **Linear Model (LM)**: A probit model for purchase incidence, and a regression model for the logarithm of the sales amount of transactions, both with $D$ and $P$ as linear covariates. The probability of a sale $\hat{b}_i$ is computed from the probit model, and expected
sales as $\hat{s}_i = E(S|B = 1, D_i, P_i) \times \hat{b}_i$, where $E(S|B = 1, D_i, P_i)$ is the expected sales for a transaction obtained from the regression model.

- **CART**: We employ the popular “Classification and Regression Tree” data-mining method with $D$ and $P$ as input variables; once for purchase incidence, and a second time for the logarithm of sales amounts of transactions. Forecasts for sales amount for all visits are then computed by multiplication in the same manner as for the linear model.

- **Survival Model (M06)**: This is an adaptation of the basic survival model proposed by Manchanda et al. (2006) for online purchase incidence. These authors employ a proportional hazards model with an advertising exposure measure as a covariate, a step function for the baseline hazard, and only observe purchases. In comparison, we employ page views as a covariate, an empirical distribution function for the baseline hazard (i.e. a Cox proportional hazards model) and employ both purchase and non-purchase visits.

- **SMP**: This is a trivariate variant of our model excluding duration, so that forecasts are constructed conditioning only on page views.

- **SMD**: This is a trivariate variant of our model excluding page views, so that forecasts are constructed conditioning only on duration.

For each method we compute the root mean square error (RMSE), both for predictions of purchase incidence and sales amount, except for M06 which only provides predictions of purchase incidence. Table 3 summarises the predictive accuracy of all methods over the holdout samples for all nine websites. The RMSE value is reported for SM, and percentage deviations from this value are reported for other methods. Positive deviations correspond to increases in RSME, and therefore lower accuracy. If the squared errors are not significantly different using a matched-pairs test at the 95% confidence level, then the percentage difference figure is given in italics.
A number of conclusions can be drawn from the results. First, our proposed model (SM) provides forecasts that are more accurate than the naïve benchmark, N, and the competing methods LM, M06 and CART. The differences are significant for both forecasts of purchase incidence and sales amount, except for CART forecasts for apple.com. Second, it is clear that duration of a visit and the number of page views are useful in forecasting sales incidence and amount, with N dominated, particularly for retailers of apparel. Third, including both $D$ and $P$ as predictors improves forecast accuracy over-and-above using either separately. Fourth, Table 3 also includes validation results for a mixture model, but we discuss these in Section 4 where it is introduced. Last, we note that small improvements in forecast accuracy for each visit have the potential to make a substantial impact because of the very large number of visits. For example, our panel of approximately 100,000 users made over 407,000 visits to amazon.com in 2007, so the total number of visits by U.S. internet-enabled households to this website alone would number in the hundreds of millions during that year.

3 Empirical Analysis

We now demonstrate the usefulness of our stochastic model for investigating the relationship between the website visit variables, purchase incidence and sales for major online retailers in three product categories: books and digital media, travel services and apparel. We showcase apple.com separately, given its unique online retail product assortment and its highly discrete pricing structure.

3.1 Books and Digital Media

The first website we examine is amazon.com, which was the world’s largest online retailer by total sales in 2007. Amazon.com sells products in a wide variety of classes, but has a particular focus on books (47% of sales) and digital media products, such as DVDs and CDs of movies and music (28% of sales). There are 407,805 visits by comScore panelists to amazon.com in our data, with 31,851 (7.81%) of these visits resulting in a purchase. Table 5
contains the page view partitions, and the number of observations in each page view range.

Table 6 reports the parameter estimates and 95% confidence intervals for the component distributions in equation (2.1), and it can be seen that all the parameters vary substantially over the page view ranges. From the estimates for the distribution \( F_2 \), the probability of a purchase being made when the number of page views is between 1 and 7 is low at \( \Pr(B = 1|1 \leq P \leq 10) = 0.6\% \). As one would expect, this increases monotonically with the number of page views, so that for visits with between 76 and 99 page views the probability of a purchase rises to \( \Pr(B = 1|76 \leq P \leq 99) = 30.9\% \). For each bivariate Gaussian copula we also compute Spearman’s rho for the dependence between sale amount and visit duration, assuming a purchase occurs. This is also reported in Table 6, and is positive for all but one page view partition, with the largest value being 0.106.

Thus, it might initially appear that dependence between the duration of a visit and the sale amount is quite low. However, Table 7 reports the matrix of pairwise marginal Spearman correlations, and that between duration and sale amount is \( \hat{\rho}_{S,D} = 0.259 \). To illustrate the usefulness of our model, Table 7 also contains the marginal Pearson sample correlations, which differ from the Spearman correlations because they do not take into account the highly non-Gaussian distribution of the variables. The Pearson correlations understate substantially the relationship between both visit variables (\( D \) and \( P \)) and the sales amount (\( S \)).

We also compute expected sales, conditional on both duration and page views. Figure 1 plots the results as a three dimensional surface “sliced” at the mid-point of each page view partition. As the number of page views increase there is a corresponding increase in the expected sale amount. However, the same is not true for duration, with there being a visit duration that results in a maximum level of expected sales for each page view level. The relationship of the two visit variables with expected sales cannot be represented in an additive fashion, and is instead a bivariate nonlinear iteration.

Figure 1 shows that the link between the two visit variables and expected sales is very
different for each variable. We examine this further by computing expected sales, marginalizing, respectively, over page views and duration. These expectations can be computed in closed form, as detailed in the Appendix. Starting with the expected sale amount and duration relationship (marginalizing over page views), denoted $E[S|D]$, Figure 2(a) shows that expected sales peak at $14.50 for visits of duration of 59 minutes. In contrast, a plot of expected sales against page views ($E[S|P]$) in Figure 2(b) shows that sales simply increase monotonically as page views increase.

Figure 2(c) graphs the expected sales conditional on duration when a purchase is made (i.e., $E[S|D, B = 1]$), showing that expected sales increase monotonically as a function of duration among just those who eventually make a purchase. Clearly, purchase incidence has a role to play. Figures 2(d) and 2(e) give the purchase probability conditional on, respectively, duration and page views. These are computed by marginalizing out the other variable as outlined in the Appendix. Figure 2(d) reveals that, for amazon.com, purchase incidence as a function of duration ($\Pr(B = 1|D)$) increases then declines, while Figure 2(e) shows that purchase incidence always increases as a function of page views. Hence, the reason for the differences between Figures 2(a) and 2(b) is that purchase incidence rises then declines as duration increases, but purchase incidence always increases as more pages are viewed. A likely reason for the effect observed in Figure 2(d) is that there is one or more segments of buyers who are goal-directed and therefore time-efficient in their purchase behavior. By constrast, members of other segments are simply browsing a website and are eventually tempted to purchase after a lengthy visit (see also, Bucklin and Sismeiro 2003 and Danaher and Mullarkey 2003). Hence, Figure 2(d) is likely due to a mixing of these broad segments. In Section 4 we show how our model can be extended to incorporate latent segmentation household-level heterogeneity.

Barnes and Noble are the second largest online book retailer in the U.S., so we also look at their website bn.com for comparison. The site offers a product range based primarily around books, in contrast to Amazon’s broader offering. In 2007 Barnes and Noble had
approximately 9% of the traffic and 6% of the total sales of the much larger retailer Amazon. Nevertheless, Table 5 shows that the partitions for the page view deciles for these two retailers are broadly similar.

Figure 2 also plots expected sale amount and probability of purchase for visits to bn.com, conditional on the website visit variables for the fitted model. In comparison to amazon.com expected sale amount peaks for slightly shorter visits of duration 52 minutes, but at a much lower value of just over $10.08. Clearly, amazon.com proves to be more successful at converting each individual visit to a higher sales amount. Moreover, if a purchase does occur, Figure 2(c) shows that expected sale amount does not increase as quickly with duration as for amazon.com. However, there appears to be little difference in the underlying purchase rates for these two websites. Figures 2(d) and 2(e) reveal that the purchase probability conditional on, respectively, duration and page views are similar for both websites. Hence, the difference in expected sale amount appears to result from higher basket totals at amazon.com. Interestingly, visits with very similar durations of 42.4 and 41.2 minutes have the maximum purchase probability for amazon.com (0.23) and bn.com (0.23), respectively.

3.2 Apparel and Travel Services

Table 1 shows that the retailers jcpenny.com, victoriassecret.com and oldnavy.com are the fifth, sixth and seventh largest retailers as measured by total number of purchases. All three are major apparel retailers, although jcpenny.com has the most diversified product lineup, victoriassecret.com is a niche retailer and oldnavy.com sells family fashion and accessories. In addition, Table 1 shows that the sites expedia.com, orbitz.com and travelocity.com are the third, fifth and seventh largest retailers as ranked by total value of sales. They all provide travel services and have a product portfolio that is more homogenous than the three apparel retailers. We estimate our stochastic model for each of these six sites and present some of the key relationships between sales and visits in Figure 3.

Figures 3(a) and 3(c) show that victoriassecret.com and jcpenny.com both derive higher sales from visits than oldnavy.com; presumably due to their differing product lineups. More
Interestingly, it appears that victoriassecret.com is particularly successful in translating visits with longer durations into higher sales amounts when a purchase is made. Moreover, all three apparel retailers appear to convert higher duration visits into higher sales more effectively than either the two book retailers or three travel service providers.

Figure 3(b) shows that the three travel service providers have differing degrees of success at converting visits of longer duration into sales amounts. Travelocity.com is most successful, with the highest expected sale amount of $141.80 for visits of duration 112 minutes. The differences between the three sites appear driven entirely by differing abilities to convert longer duration visits into purchase events. Once the expected sale amount is computed, conditional on a purchase being made, there is minimal difference between the three travel service providers as depicted in Figure 3(d), which likely reflects the similarity of the products offered at these travel websites.

Figures 3(e) and 3(f) depict the probability of a sale being made against the number of page views for the six websites. Generally, higher page views correspond to a higher probability of purchase. Interestingly, the sites that are least successful at converting longer duration visits into sales, are not necessarily poor at converting more page views into more purchases. For example, oldnavy.com has the highest purchase probability among all retailers, as the number of page views increases.

### 3.3 Research Online and Buy Online Later

A final observation that applies just to the apparel retailers concerns Figure 3(c). A subtle feature of this plot is a small dip in sales between 5 and 10 minutes. This is due to some short duration visits (less than 10 mins) where the expected sales are relatively high. We conjecture that this is due to recent prior visits to the website that are strictly browsing, without a sale being made. During this time a shopper likely peruses the merchandise, possibly checking out competing websites. Eventually when the decision to purchase is made, the transaction time is relatively quick, because product research has been completed prior to the actual purchase visit. The likelihood of such behavior is very plausible because online
research prior to bricks-and-mortar purchase is commonplace (Krillion 2008; Mendelsohn et al. 2006). All we are suggesting here is the eventual purchase is made online rather than in-store.

We test this conjecture by dividing purchase visits into those that are fast ($\leq 10$ mins) and slow ($> 10$ mins), and then calculate the proportion of households in these two groups that have visited the same website within 48 hours prior to the eventual purchase visit, but have not purchased anything during those prior visits. Table 8 reports these proportions. Fast purchasers of apparel products research online before making the purchase much more often than slow purchasers (between 27.5% more often for jcpenney.com and 62.7% more often for oldnavy.com), consistent with our conjecture. This behavior is replicated by purchasers of travel services, but to a lesser extent. However, there is very little difference between the online product research undertaken by fast and slow purchasers of books or digital media products, where the purchase risk is lower.

### 3.4 Discrete Sales Categories for Apple.com

Apple.com has a high rate of conversion of visits into purchases at 29.1% in 2007. Moreover, the estimated relationship between website visit and sales variables is very different to that of the other retailers in our study. Figure 4 shows the probability of purchase against duration in Figure 4(a) and number of page views in Figure 4(b). Apple.com visitors have much fewer page views on average than other online retailers, but these convert much more rapidly into higher purchase probabilities compared with retailers in our study. The expected sale amount is low because sales are predominantly for a single song. For visits where a purchase is made, Figures 4(c) and Figure 4(d) plot the expected sale amount against duration and page views, respectively. The expected sale amount peaks strongly at visits of duration 4 minutes. Compare this to the relationship for retailers of books, apparel and travel services (Figures 2(c), 3(c) and 3(d)), where purchasing visits have a higher expected sale amount as duration increases. For visits with a small number of pages views, the expected sale amount at apple.com is close to $0.99. However, this increases markedly for visits with 10 or more
We also compute the Spearman’s correlations between the variables for the fitted stochastic model. There is a lower dependence between the visit and sales variables when compared with amazon.com in Table 7. This is particularly true for the purchase indicator, with \( \hat{\rho}_{B,D} = 0.223 \) and \( \hat{\rho}_{B,P} = 0.183 \). This suggests that visitors to apple.com are more goal-directed than those at amazon.com, as might be anticipated for a website that is tailored primarily towards transactions rather than browsing.

4 Latent Segmentation

4.1 Bayesian Model

While both the NBD and Inverse Gaussian are popular model choices for data that arise from heterogeneous populations, it is likely the data exhibit further heterogeneity. For example, Moe and Fader (2004) propose a taxonomy of four groups of online shoppers, based on visit behavior and purchase incidence. Therefore, to further account for household-level consumer heterogeneity we employ a finite mixture model with latent segmentation. This approach is well-established in marketing (Kamakura and Russell 1989; Allenby and Rossi 1998), although usually using a mixture of normals, whereas we consider a mixture of the quad-variate stochastic models at equation (2.1).

Consider a mixture with \( M \) latent segments, and probability \( \pi_l \) that a household is a member of segment \( l \), then the distribution of the visit and sales variables is

\[
F(S, B, D, P) = \sum_{l=1}^{M} \pi_l F^l(S, B, D, P) = \sum_{l=1}^{M} \pi_l \left( F_1^l(S, D|B, P)F_2^l(B|P)F_3^l(P) \right),
\]

(4.1)

where segments are denoted with a superscript.

Bayesian estimation of mixture models is popular, with posterior inference computed using Markov chain Monte Carlo (MCMC) methods; for example, see Diebolt and Robert (1994). This includes the ability to profile the segments, which can be difficult using other likelihood-
based methods (Wedel and DeSarbo 2002). Latent multinomial variables are introduced to specify segment membership for each household \( h \), where \( L_h = l \) if household \( h \) is a member of segment \( l \). We denote the set of latent variables for all \( H \) households as \( L = \{L_1, \ldots, L_H\} \). Conditional on \( L \), the likelihood of the mixture model is simplified with respect to the parameters of segment \( F^l \). MCMC samplers that generate the latent variables explicitly, and the parameters conditional upon these, are popular; for example, see Diebolt and Robert (1994) and Lenk and DeSarbo (2000). In the Supplementary Material we outline such a sampling scheme to generate \( J \) iterates from the posterior distribution of the mixture model parameters, augmented with the household latent variables.

In our Bayesian analysis we adopt a Dirichlet prior for \( \pi = (\pi_1, \ldots, \pi_M) \sim \text{Dirichlet}(\alpha) \), which is a common choice because it has the Bayesian property of being conjugate to the multinomial. To make the mixture model more flexible, we make \( \alpha \) a hyperparameter with a uniform hyperprior on \([0, 2]^M\), which ensures the prior on \( \pi \) is flat at the prior expected value of \( E(\alpha) = 1 \). The priors on the parameters of the component distributions \( F^1_l, F^2_l, F^3_l \) are the same across segments and proper, which is important to facilitate model selection, but noninformative, so that the posterior distribution of the parameters is dominated by the likelihood. We select the number of segments using a deviance information criterion (DIC), suggested by Celeux et al. (2006), which is more appropriate for model choice in mixture models than that introduced by Spiegelhalter et al. (2002). Adopting the notation of Frühwirth-Schnatter and Pyne (2010), this is defined for each value of \( M \) as

\[
\text{DIC}_{4\alpha}(M) = -4E_{\Phi} \left[ \log \mathcal{L}(\Phi; y) | y \right] + 2 \log \mathcal{L}(\hat{\Phi}; y) + 2E_{\Phi} \left[ \text{EN}(\Phi) | y \right].
\]

In the above, \( y \) denotes the data, \( \Phi \) the parameters of all \( M \) segments, \( \hat{\Phi} \) is the posterior mode, \( \mathcal{L} \) is the likelihood, EN is the entropy and the expectations are with respect to the posterior distribution. This criterion is more robust to label-switching; see Lenk and DeSarbo (2000) and Stephens (2000) for discussions of this problem. The likelihood \( \mathcal{L} \) is from equation (4.1),
and the criterion can be computed efficiently using the output of the sampling scheme as discussed in Frühwirth-Schnatter and Pyne (2010; Sec. 3.3). We compute $\text{DIC}_{\text{da}}(M)$ for $M = 1, 2, 3, 4$, and select $M$ that maximizes its value. We found that for all websites $M = 4$ was the optimal value, except for oldnavy.com, which had $M = 3$ segments.

### 4.2 Predictions

Parameter estimates and other inference can be readily computed using the output from the sampling scheme. We estimate the mixture model using the calibration data $y$, and then forecast purchase incidence and sale amount using the Bayesian predictive mass and expectation $\Pr(B_i = 1|D_i, P_i, y)$ and $E(S_i|D_i, P_i, y)$ for each visit $i$ in the holdout sample of the validation study. These differ depending on whether, or not, the household $h$ associated with visit $i$ also had visits recorded in the calibration data. Nevertheless, they can be computed using the MCMC output in both cases, as outlined in the Supplementary Materials. Table 3 includes the performance of these forecasts in the MIX column. For amazon.com and victoriassecret.com there is a significant increase in forecast accuracy. For the remaining sites forecasts are either not meaningfully improved, or slightly less accurate for sales amounts at the three travel sites.

### 4.3 Segmentation for Oldnavy.com

Also of interest are the profiles of the market segments. In Bayesian estimation these are computed with the parameters and latent variables integrated out with respect to the posterior distribution, rather than conditional on their point estimates. For example, if $\Phi^l$ denotes the parameters of the $l$th segment, then the mean of the $l$th fitted segment is

$$E(S, B, D, P|L_h = l, y) = \int E(S, B, D, P|L_h = l, \Phi^l) f(\Phi^l|y) d\Phi^l \approx \frac{1}{J} \sum_{j=1}^{J} (S, B, D, P)^{[j]}.$$

Here, each iterate $(S, B, D, P)^{[j]} \sim f(S, B, D, P|L_h = l, \Phi^l, [j])$ is generated from segment $l$ with parameter values $\Phi^l, [j] \sim f(\Phi^l|y)$ obtained at sweep $j$ of the MCMC sampling scheme.
Estimates of other moments or distributional summaries for each segment can also be computed in a similar fashion.

To illustrate, Table 9 provides profiles for each of the three segments of oldnavy.com obtained using the calibration data. The top portion of the table reports the expectations of the four variables, and also the two ratios $P/D$ and $S/P$. The first ratio is a measure of how fast a visitor progresses through the site (i.e., search velocity), while the second ratio is a measure of a visitor’s expected spend in response to page exposure. Households in segment 1 have low purchase incidence, low expected spend and progress through the pages with the highest search velocity. In comparison, households in segment 3 have the highest purchase incidence, highest expected spend and highest expected spend per page. Moe (2003) and Moe and Fader (2004) characterise online shoppers into four groups: directed buyers, deliberation visitors, hedonic browsers and knowledge-building visitors. For our oldnavy.com segmentation, Segment 1 has characteristics consistent with those of knowledge-building behavior, while households in Segment 3 exhibit goal-directed behavior that is more consistent with directed buyers and deliberation visitors. Segment 2 households have characteristics consistent with hedonic browsers.

The posterior probability that a specific household $h$ is in segment $l$ is

$$
\Pr(L_h = l | y) = \int \Pr(L_h = l | \pi, \Phi, y) f(\pi, \Phi | y) d(\pi, \Phi) 
\approx \frac{1}{J} \sum_{j=1}^{J} \Pr(L_h = l | \pi^{[j]}, \Phi^{[j]}, y) = \hat{\omega}_h^l.
$$

(4.2)

Here, $\{\pi^{[j]}, \Phi^{[j]}\}_{j=1}^{J}$ are the Monte Carlo iterates output from the MCMC scheme, and the Supplementary Material shows how to compute the conditional probability in equation (4.2). The estimates $\hat{\omega}_h^l$ differ for each household in the sample and should not be confused with an estimate of the probability $\pi_l$ in equation (4.1), which is not household specific.

To see if the behavior of households in each segment extends beyond their activity at oldnavy.com, we also construct three general internet activity variables for each household
using the comScore transaction and session data. These are the total online spend, total number of online transactions and total number of sessions at the top 100 websites across the entire year of 2007. Using the oldnavy.com data, for each household $h$ we also compute the posterior probability of membership of each segment, $\hat{\omega}_h^j$, and allocate each observation to the segment with the highest probability. The bottom row of Table 9 reports the number of households that are allocated to each segment, and the middle portion reports the sample means of the three general internet activity variables. These show that the consumer behavior identified in our latent segmentation using the oldnavy.com data, extends into similar general internet activity at a household level.

5 Conclusion

In this research we develop a stochastic model for website visit duration, page views, purchase incidence and sales amount. Previous work has modeled the bivariate distributions of visit duration and purchase incidence (Lin et al. 2010; Moe and Fader 2004; Montgomery et al. 2004; Van den Poel and Buckinx 2005), and visit duration and sales (Danaher and Smith 2011). However, ours is the first study to simultaneously handle all four of these key elements of online browsing and purchasing.

From a managerial perspective, we show that the two stickiness measures are important indicators of whether a sale will occur, and for the amount of the sale. This is consistent with an earlier empirical result by Montgomery et al. (2004), who were able to predict eventual purchase incidence with 40% accuracy using information from just the first 6 pages of a website visit. Interestingly, for books, digital media and travel service websites there is a value of duration that maximizes expected sales, while expected sales generally increases monotonically with page views. Managers will also be interested to learn that while much attention has been devoted to research online, buy offline (e.g., Thackston 2009), there is a parallel phenomenon of research online prior to purchasing (also) online. Such situations are flagged by prior visits to a website, but the eventual purchase is comparatively quick.
in a subsequent visit. Therefore, online retailers should not necessarily be discouraged by the high proportion of non-sale visits (Moe and Fader 2004; Venkatesh and Agarwal 2006). We found prior online research to be especially prevalent for apparel and travel products, no doubt because such categories entail more involved purchases, and the monetary amounts are higher, than for books, DVDs and songs. We find that although websites within the same product category have different expected sales as a function of duration and page views, the relationship of these two visit variables with purchase incidence is similar. Therefore, it seems most likely that the difference in sales amounts is due to the product offerings. Our latent class segmentation for oldnavy.com identified three distinct market segments. The smallest segment consists of more engaged customers who exhibit greater goal-directed behavior. Members of the largest segment exhibit browsing behavior, while those of the remaining segment are very low spending customers. Such segmentation is consistent with the classification discussed by Moe and Fader (2004), while profiling of the wider online activities of these households suggests that this behavior extends beyond their visits to oldnavy.com.

On the methodological front, we propose a quad-variate distribution for the website visit and sales variables that is constructed by composition from carefully selected components that fit well empirically. The framework has a number of practical advantages. First, estimation is fast, so that the approach is practical given the very large size of the data that arise in studies of online retail behavior. Second, expectations of the sales variables, conditional on either visit variable separately, or both together, can be computed from the stochastic model without reverting to bivariate and trivariate numerical integration. Third, computing these expectations from the quad-variate distribution, rather than modeling them directly in a regression style framework, accounts for the simultaneous determination of all four variables. Fourth, the model can extended to a finite mixture of the quad-variate distributions, so that consumer heterogeneity can be accounted for at the household level. Last, we adapt the model to cope with the discrete pricing used by retailers such as apple.com. To achieve
this we exploit the flexibility of the bivariate copula component to model a combination of continuous and discrete marginals; something that would otherwise be difficult.

Our validation exercise shows that the proposed model outperforms a number of alternative approaches for the websites examined. This indicates that the nonlinear dependence between the variables is better captured by the model, and that the two website stickiness measures provide valuable information when predicting purchase incidence and sales. Importantly, our model is parametric and easy to extend into a variety of directions. Future work could include adding another layer to the model to incorporate website-level covariates, as used by Bucklin and Sismeiro (2003) and Danaher et al. (2006). This is straightforward by making the parameters of the component distributions functions of the covariates, with estimation undertaken using maximum likelihood in the same manner as outlined. By using time-based covariates, this also allows for the modeling of dynamic behaviour similar to that examined by Moe and Fader (2004). In addition, following Manchanda et al. (2006), a possible further extension is a hierarchical model that incorporates household-level heterogeneity. This would provide an alternative to latent class segmentation.

Appendix: Sales Summaries

In this appendix we show how to use the stochastic model in Section 2.2 to compute several summary measures of sales. These include both the probability of a purchase and the expected spend, both conditional upon only one visit variable and marginalized over the second. Page views can be marginalized over as follows:

\[
\Pr(B = 1|D) = \frac{\sum_{P=1,2,...} f(P, B = 1, D)}{\sum_{P=1,2,...} f(D)} = \frac{\sum_{P=1,2,...} f(P, B = 1, D)}{\sum_{P=1,2,...} \sum_{B=0,1} f(P, B, D)}
\]

Because the page view variable is partitioned into \(K\) ranges in our stochastic model, the
summations in $P$ can be replaced with summations over the $K$ partitions, so that

$$\Pr(B = 1|D) = \frac{\sum_{k=1}^{K} f_{1D}(D|B = 1, P \in \tilde{P}_k)\Pr(B = 1|P \in \tilde{P}_k)\Pr(P \in \tilde{P}_k)}{\sum_{k=1}^{K} \sum_{B=0,1} f_{1D}(D|B, P \in \tilde{P}_k)\Pr(B|P \in \tilde{P}_k)\Pr(P \in \tilde{P}_k)}.$$  \hspace{1cm} (A1.1)

The probability $\Pr(P \in \tilde{P}_k)$ is computed from the NBD by simple summation over the mass function, and all other components in the expression above are known from the model definition, so that it can be computed analytically. Derivation of the expected spend conditional on duration only is similar. We first note that $E(S|D) = \sum_{B=0,1} \sum_{P=1,2,...} E(S|D, P, B)\Pr(P, B|D)$, but because $S = 0$ when $B = 0$, this expression simplifies to

$$E(S|D) = \sum_{P=1,2,...} E(S|D, P, B = 1)\Pr(P, B = 1|D).$$

The rest of the derivation follows the same expansion of $\Pr(P, B = 1|D)$ as above, and the substitution of the summations in $P$ with those over the partition, to give

$$E(S|D) = \frac{\sum_{k=1}^{K} E(S|D, B = 1, P \in \tilde{P}_k)f_{1D}(B|B = 1, P \in \tilde{P}_k)\Pr(B = 1|P \in \tilde{P}_k)\Pr(P \in \tilde{P}_k)}{\sum_{k=1}^{K} \sum_{B=0,1} f_{1D}(D|B, P \in \tilde{P}_k)\Pr(B|P \in \tilde{P}_k)\Pr(P \in \tilde{P}_k)}.$$  \hspace{1cm} (A1.2)

The expected spend conditional upon duration and that a purchase is made is obtained from the above by noting that $E(S|D) = E(S|B = 1, D)\Pr(B = 1|D)$ and rearranging to get

$$E(S|B = 1, D) = \frac{E(S|D)}{\Pr(B = 1|D)},$$  \hspace{1cm} (A1.3)

where the numerator and denominator are given above. Computation of the terms at equations (A1.1)–(A1.3) is easily undertaken for a range of values for $D$. 

27
References

Allenby, Greg M. and Peter E. Rossi (1998), “Marketing models of consumer heterogeneity”, *Journal of Econometrics*, 89, 1-2, 57-78.

Bucklin, Randolph E. and Catarina Sismeiro (2003), “A Model of Web Site Browsing Behavior Estimated on Clickstream Data”, *Journal of Marketing Research*, 40 (August), 249-267.

Celeux, G., M. Hurn and C. Robert (2000), “Computational and inferential difficulties with mixture posterior distributions”, *Journal of the American Statistical Association*, 95, 957-970.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004), *Copula Methods in Finance*, New York, NY: Wiley.

Danaher, Peter J. (2007), “Modeling Page Views Across Multiple Websites With An Application to Internet Reach and Frequency Prediction”, *Marketing Science*, 26, 3 (May/June), 422-437.

Danaher, P.J. and Mullarkey, G. (2003), “Factors Affecting Online Advertising Recall: A Study of Students”, *Journal of Advertising Research*, 43, 3 (September), 252-267.

Danaher, Peter J., Mullarkey, G and Essegaier, S. (2006), “Factors Affecting Website Visit Duration: A Cross-Domain Analysis”, *Journal of Marketing Research*, 43 (May), 182-194.

Danaher, Peter J. and Michael S. Smith (2011), “Modeling Multivariate Distributions Using Copulas: Applications in Marketing”, (with discussion and rejoinder), *Marketing Science*, 30, 4-21.

Diebolt, Jean and Christian P. Robert (1994), “Estimation of Finite Mixture Distributions through Bayesian Sampling”, *Journal of the Royal Statistical Society*, Series B, 56, 2, 363-375.

Früwirth-Schnatter, S. and Pyne, S. (2010), “Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions”, *Biostatistics*, 11, 2, 317-336.

Glady, Nicolas, Aurelie Lemmens and Christophe Croux (2010), “Modeling Within- and Across-Customer Association in Lifetime Value with Copulas”, CentER Discussion Paper.

Hougaard, Philip (1984), “Life table methods for heterogeneous populations: Distributions describing the heterogeneity”, *Biometrika*, 71, 1, 75-83.

Huang, C.-Y. and C.-S. Lin (2006), “Modeling the Audience’s Banner Ad Exposure for Internet Advertising Planning”, *Journal of Advertising*, 35, 2, 23-37.
Joe, Harry (2006), “Asymptotic efficiency of the two-stage estimation method for copula-based models”, *Journal of Multivariate Analysis*, 94, 2, 401-419.

Johnson, Eric J., Steven Bellman and Gerald L. Lohse (2003), “What Makes a Website Sticky? Cognitive Lock-In and the Power Law of Practice”, *Journal of Marketing*, 67, 2 (April), 62-75.

Johnson, Norman L., Samuel Kotz and N. Balakrishnan (1994) *Continuous Univariate Distributions*, 2nd Ed., Vol.1, Wiley.

Kamakura, Wagner, A. and Gary J. Russell (1989), “A Probabilistic Choice Model for Market Segmentation and Elasticity Structure”, *Journal of Marketing Research*, 26, 4, 379-390.

Krillion (2008), “New Survey Finds 67 Percent of Shoppers Invest 30+ Percent of Their Total Shopping Time Researching What To Buy”, http://www.krillion.com/xAV-news-20080324 _retailing_survey, Accessed 26 October 2010.

Lenk, Peter J. and Wayne S. DeSarbo (2000), “Bayesian Inference for Finite Mixtures of Generalized Linear Models with Random Effects”, *Psychometrika*, 65, 1, 93-119.

Lin, Lin, Paul Jen-Hwa Hu, Olivia R. Liu Sheng and Johnny Lee (2010), “Is Stickiness Profitable for Electronic Retailers?” *Communications of the ACM*, 53, 3 (March), 132-136.

Manchanda, Puneet, Jean-Pierre Dube, Khim Yong Goh and Pradeep K. Chintagunta (2006), “The Effect of Banner Advertising on Internet Purchasing”, *Journal of Marketing Research*, 43 (February), 98-108.

Mendelsohn, T., C.A. Johnson and S. Meyer (2006), “Understanding U.S. Cross-Channel Shoppers”, Forrester Research, April 19.

Moe, Wendy W. (2003), “Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream”, *Journal of Consumer Psychology*, 13, (1&2), 29-40.

Moe, Wendy W. and Peter S. Fader (2004), “Dynamic Conversion Behavior at e-Commerce Sites”, *Management Science*, 50, 3, 326-335.

Montgomery, Alan L, Shibo Li, Kannan Srinivasan and John C. Liechty (2004), “Modeling Online Browsing and Path Analysis Using Clickstream Data”, *Marketing Science*, 23, 4, 579-595.

McNeil, A. J., R. Frey and R. Embrechts, (2005), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princton: NJ.

Nelsen, R., (2006), *An Introduction to Copulas*, 2nd ed., Springer.
Oyer, Paul (2000), “A Theory of Sales Quotas with Limited Liability and Rent Sharing”, *Journal of Labor Economics*, 18, 3, 405-426.

Park, Sungho and Sachin Gupta (2012), “Handling Endogenous Regressors by Joint Estimation Using Copulas”, *Marketing Science*, 31, 4, 567-586.

Shih, J.H. and T.A. Louis (1995), “Inferences on the Association Parameter in Copula Models for Bivariate Survival Data”, *Biometrics*, 51, 1384–1399.

Song, Peter, (2000), “Multivariate Dispersion Models Generated from Gaussian Copula”, *Scandinavian Journal of Statistics*, 27, 305-320.

Spiegelhalter, D., N. Best., B. Carlin and A. Van Der Linde, (2002), “Bayesian measures of model complexity and fit”, *Journal of the Royal Statistical Society*, Series B, 64, 583-639.

Stephens, Matthew, (2000), “Dealing with label switching in mixture models”, *Journal of the Royal Statistical Society*, Series B, 62(4), 795-809.

Thackston, Karon (2009), “Cross-Channel Shoppers: Are ROBO Shoppers Faking Out Your PPC Conversion Rate?”, http://www.businessknowhow.com/internet/crosschannel.htm, accessed 5 August 2010.

Trivedi, P. and D. Zimmer, (2005), “Copula Modeling: An Introduction for Practitioners”, *Foundations and Trends in Econometrics*, 1, 1, 1-110.

Van den Poel, Dirk and Wouter Buckinx (2005), “Predicting Online Purchasing Behavior”, *European Journal of Operational Research*, 166, 557-575.

Venkatesh, Viswanath and Ritu Agarwal (2006), “Turning Visitors into Customers: A Usability-Centric Perspective on Purchase Behavior in Electronic Channels”, *Management Science*, 52, 3, 367-382.

Wedel, Michel and Wayne S. DeSarbo (2002), “Finite Segment Derivation and Profiling Via a Finite Mixture Model Framework’, *Marketing Letters*, 13, 1, 17-25.
Table 1: Top ranking online retailers in the 2007 ComScore panel. The retailers are ranked by both total sales (in $1000) over the panel and by total number of purchases, expressed as a percentage of the total number of purchases observed from the panel.

Figure 1: Expected sale amount at amazon.com as a function of duration of visit and number of page views (on the logarithmic scale) resulting from the parametric model. Note that the impact of duration and page views on expected sales is not additive in form.
| Component | Distribution       | Parameters | Density/Probability Mass                                                | Domain  |
|-----------|--------------------|------------|----------------------------------------------------------------------|---------|
| $F_3(P)$  | Truncated NBD      | $r, q$     | $\Pr(P = p) = \frac{g(p)}{1-g(0)}$, with $g(p) = \left(\frac{r + p - 1}{p}\right) q^r (1 - q)^p$ | $p = 1, 2, \ldots$ |
| $F_2(B|P)$| Bernoulli          | $p_B$      | $f_2(b) = p_B^b (1 - p_B)^{1-b}$                                      | $b \in \{0, 1\}$ |
| $F_{1D}(D|B = 0, P)$| Inverse Gaussian  | $\mu_0, \lambda_0$ | $f_{1D}(d) = \sqrt{\frac{2\mu_0}{\lambda_0}} \exp \left\{-\frac{\lambda_0}{2\mu_0} (d - \mu_0)^2\right\}$ | $d > 0$ |
| $F_{1D}(D|B = 1, P)$| Inverse Gaussian  | $\mu_1, \lambda_1$ | $f_{1D}(d) = \sqrt{\frac{2\mu_1}{\lambda_1}} \exp \left\{-\frac{\lambda_1}{2\mu_1} (d - \mu_1)^2\right\}$ | $d > 0$ |
| $F_{1S}(S|B = 1, P)$| Log-logistic      | $\mu_S, \sigma_S$ | $f_{1S}(s) = \exp\left(\frac{\text{log}(s) - \mu_S}{\sigma_S}\right)$ | $s > 0$ |
| $C$       | Gaussian Copula    | $\theta$   | $c(u, v; \theta) = (1 - \theta^2)^{-1/2} \exp \left\{-\theta^2 (w_u^2 + w_v^2 - 2\theta w_u w_v)\right\}$ | $(u, v) \in [0, 1]^2$ |

Table 2: Component distributions in the joint distribution of $(S, B, D, P)$ in Section 2. Apart from the truncated NBD, the parameters vary over the $K$ page view partitions. For the Gaussian copula $w_u = \Phi^{-1}(u)$ and $w_v = \Phi^{-1}(v)$, with $\Phi$ the standard normal distribution function. Note that the Log-Logistic distribution is replaced by the Empirical Distribution Function in the case of apple.com to account for highly discrete pricing.
Table 3: Predictive performance of the different methods for all nine websites in the validation study, including the mixture model in Section 4 (MIX). The root mean square error (RMSE) of forecasts for purchase incidence $B$, and also sales amount $S$, are reported for method SM (our proposed stochastic model). For the other methods, the percentage difference between the RMSE values for each method, and that for SM, is reported. Positive percentages correspond to an increase in RMSE compared to method SM, and therefore lower predictive accuracy. When these are not significantly different from zero, then the percentage figure it is given in italics. A mixture model was not considered for the website apple.com, so that no value is reported for this case.
| Website                  | Non-Purchases | Purchases | Total   | Households | $K$ |
|-------------------------|---------------|-----------|---------|------------|-----|
| **Books & Digital Media**|               |           |         |            |     |
| amazon.com              | 375954        | 31851     | 407805  | 53929      | 26  |
| barnesandnoble.com      | 34993         | 2923      | 37916   | 13037      | 18  |
| apple.com               | 191039        | 77398     | 268437  | 48898      | 12  |
| **Apparel**             |               |           |         |            |     |
| oldnavy.com             | 56128         | 2855      | 58983   | 12876      | 25  |
| jcpenney.com            | 106939        | 7336      | 114275  | 26443      | 24  |
| victoriassecret.com     | 57598         | 5014      | 62612   | 15308      | 17  |
| **Travel Services**     |               |           |         |            |     |
| expedia.com             | 138263        | 3649      | 141912  | 32671      | 25  |
| orbitz.com              | 80558         | 2753      | 83311   | 25522      | 19  |
| travelocity.com         | 77833         | 2516      | 80349   | 22574      | 25  |

Table 4: Sample sizes for the nine websites in our study. Each observation corresponds to a visit, and the number of visits that ultimately result in a purchase are given, along with those that do not. The fifth column reports the number of households from which these visits originated. The last column contains the optimal number of page view partitions identified by cross-validation for each website, as discussed in Section 2.3.

| $k$ | $\hat{P}_k$ | No Sale | Purchases | $\hat{P}_k$ | No Sale | Purchases |
|-----|-------------|---------|-----------|-------------|---------|-----------|
| 1 & 2 | 1-7, 7-9  | 244266  | 2523      | 1-10, 11-13 | 26261  | 394       |
| 3 & 4 | 10-11, 12-12 | 33360  | 2732      | 14-15, 16-16 | 1879  | 283       |
| 5 & 6 | 13-13, 14-14 | 15163  | 2105      | 17-18, 19-20 | 1699  | 370       |
| 7 & 8 | 15-16, 17-17 | 17009  | 2964      | 21-22, 23-23 | 935   | 254       |
| 9 & 10 | 18-18, 19-20 | 12481  | 2640      | 24-26, 27-29 | 1260  | 373       |
| 11 & 12 | 21-21, 22-23 | 9168   | 2348      | 30-32, 33-35 | 795   | 281       |
| 13 & 14 | 24-24, 25-26 | 7296   | 2045      | 36-40, 41-46 | 848   | 326       |
| 15 & 16 | 27-28, 29-30 | 7126   | 2316      | 47-54, 55-66 | 681   | 319       |
| 17 & 18 | 31-33, 34-35 | 6368   | 2395      | 67-92, 93-500 | 635   | 323       |
| 19 & 20 | 36-39, 40-43 | 6924   | 2747      |             |        |           |
| 21 & 22 | 44-47, 48-54 | 5811   | 2295      |             |        |           |
| 23 & 24 | 55-62, 63-75 | 5313   | 2353      |             |        |           |
| 25 & 26 | 76-99, 100-500 | 5651   | 2388      |             |        |           |
| **Total** |            | 375954  | 31851     | 34993      | 2923   |           |

Table 5: Optimal page view partitions for the two book retailers, presented as two adjacent partitions per row. The sample sizes in the paired partitions are reported, broken down by observations where no sale was made, and those visits where one or more items were purchased.
| Page Views $(P)$ | Copula Dependence $\hat{\theta}$ | $\hat{\rho}^C$ | $F_{1,S}(S|B=1,P)$ $\hat{\mu}_S$ $\hat{\sigma}_S$ | $F_{1,D}(D|B=1,P)$ $\hat{\lambda}_1$ | $F_{1,D}(D|B=0,P)$ $\hat{\mu}_0$ $\hat{\lambda}_0$ | $F_2(B|P)$ $\hat{p}_B$ |
|-----------------|--------------|---------|-----------------|-----------------|-----------------|-----------|
| $1 \leq P \leq 7$ | 0.085 0.081 | 3.24 0.481 | 4.76 5.21 | 4.41 2.98 | 0.006 |
| | (0.018, 0.15) | (3.19, 3.28) (0.46, 0.504) | (4.51, 5.00) (4.81, 5.61) | (4.39, 4.43) (2.97, 3.0) | (0.006, 0.007) |
| $10 \leq P \leq 11$ | 0.064 0.061 | 3.23 0.477 | 8.54 13.6 | 10.1 9.38 | 0.067 |
| | (0.013, 0.115) | (3.19, 3.27) (0.459, 0.496) | (8.22, 8.86) (12.7, 14.5) | (10.0, 10.3) (9.2, 9.5) | (0.064, 0.07) |
| $13 \leq P \leq 16$ | -0.001 -0.001 | 3.28 0.469 | 10.2 21.1 | 12.3 12.9 | 0.118 |
| | (-0.062, 0.061) | (3.23, 3.33) (0.446, 0.493) | (9.80, 10.7) (19.3, 22.9) | (12.0, 12.5) (12.5, 13.3) | (0.111, 0.124) |
| $15 \leq P \leq 16$ | 0.053 0.051 | 3.31 0.487 | 12.3 27.9 | 14.5 16.8 | 0.143 |
| | (0.01, 0.096) | (3.27, 3.35) (0.469, 0.505) | (11.9, 12.7) (26.2, 29.7) | (14.3, 14.7) (16.3, 17.2) | (0.138, 0.149) |
| $18 \leq P \leq 18$ | 0.078 0.075 | 3.39 0.496 | 15.5 35.1 | 16.3 20.1 | 0.164 |
| | (0.015, 0.141) | (3.34, 3.45) (0.47, 0.523) | (14.8, 16.1) (31.9, 38.3) | (15.9, 16.7) (19.2, 20.9) | (0.155, 0.174) |
| $21 \leq P \leq 21$ | 0.111 0.106 | 3.42 0.481 | 17.5 46.6 | 18.7 25.3 | 0.184 |
| | (0.043, 0.178) | (3.36, 3.48) (0.454, 0.511) | (16.7, 18.2) (41.9, 51.2) | (18.2, 19.3) (24.1, 26.4) | (0.173, 0.196) |
| $24 \leq P \leq 24$ | 0.065 0.062 | 3.48 0.498 | 19.4 55.2 | 21.1 31.6 | 0.209 |
| | (-0.004, 0.134) | (3.42, 3.54) (0.468, 0.53) | (18.6, 20.3) (49.4, 61.0) | (20.4, 21.7) (29.9, 33.3) | (0.195, 0.223) |
| $27 \leq P \leq 28$ | 0.054 0.051 | 3.55 0.483 | 22.9 69.8 | 23.0 37.4 | 0.230 |
| | (-0.002, 0.109) | (3.50, 3.60) (0.461, 0.507) | (22.2, 23.7) (64.2, 75.4) | (22.5, 23.6) (35.7, 39.1) | (0.227, 0.251) |
| $31 \leq P \leq 33$ | 0.043 0.041 | 3.56 0.503 | 26.0 79.9 | 26.0 42.5 | 0.268 |
| | (-0.007, 0.092) | (3.52, 3.60) (0.482, 0.524) | (25.3, 26.8) (74.2, 85.7) | (25.4, 26.6) (40.7, 44.4) | (0.256, 0.279) |
| $36 \leq P \leq 39$ | 0.013 0.013 | 3.60 0.519 | 28.8 104.7 | 29.5 56.4 | 0.281 |
| | (-0.036, 0.062) | (3.55, 3.64) (0.498, 0.542) | (28.0, 29.5) (97.2, 112.2) | (28.8, 30.2) (54.0, 59.0) | (0.269, 0.293) |
| $44 \leq P \leq 47$ | 0.071 0.068 | 3.62 0.544 | 35.5 119.7 | 34.4 65.5 | 0.273 |
| | (0.011, 0.131) | (3.56, 3.68) (0.516, 0.574) | (34.2, 36.7) (108.9, 130.5) | (33.4, 35.3) (61.9, 69.1) | (0.258, 0.288) |
| $55 \leq P \leq 62$ | 0.049 0.047 | 3.71 0.559 | 43.8 178.0 | 42.7 88.0 | 0.297 |
| | (-0.007, 0.106) | (3.65, 3.77) (0.532, 0.587) | (42.5, 45.1) (163.3, 192.7) | (41.5, 43.8) (83.3, 92.7) | (0.283, 0.312) |
| $76 \leq P \leq 99$ | 0.079 0.075 | 3.79 0.580 | 59.7 224.7 | 55.0 143.0 | 0.309 |
| | (0.023, 0.134) | (3.73, 3.85) (0.553, 0.608) | (58.0, 61.5) (206.5, 242.7) | (53.7, 56.3) (135.3, 150.7) | (0.294, 0.323) |

Parameters of the NBD $F_3$: $\hat{r} = 0.7660 \pm 4.27 \times 10^{-6}$; $\hat{q} = 0.0600 \pm 2.97 \times 10^{-7}$; $E(P) = 13.56$; Std. Dev. $(P) = 15.033$

Table 6: Estimates of the parameters of the stochastic model for amazon.com, with 95% confidence intervals given below in parentheses. To save space, only estimates for the 13 even numbered partitions are given.
| S | B | D | P |
|---|---|---|---|
| **Spearman Correlations** | | | |
| S | 1 | 0.9984 | 0.2589 | 0.3236 |
| B | 1 | 0.2574 | 0.3226 | |
| D | 1 | | 0.6761 | |
| P | | 1 | | |
| **Pearson Sample Correlations** | | | |
| S | 1 | 0.4161 | 0.1324 | 0.1653 |
| B | 1 | 0.2244 | 0.3034 | |
| D | 1 | | 0.7218 | |
| P | | 1 | | |

Table 7: Marginal pairwise Spearman dependence measures from the fitted model for amazon.com, and the Pearson sample correlations. The high Spearman dependence between $S$ and $B$ is due to the very strong zero-inflation of the sales data, which is commonplace in online retail.

| Retail Website        | Fast Buys | Slow Buys | % Difference |
|-----------------------|-----------|-----------|--------------|
| **Books & Digital Media** | | | |
| amazon.com            | 34.69%    | 34.23%    | 1.3%         |
| barnesandnoble.com    | 24.62%    | 22.03%    | 11.8%        |
| apple.com             | 33.33%    | 35.19%    | -5.3%        |
| **Apparel**           | | | |
| oldnavy.com           | 51.71%    | 31.78%    | 62.7%        |
| jcpenny.com           | 37.99%    | 29.79%    | 27.5%        |
| victoriassecret.com   | 36.93%    | 26.25%    | 40.7%        |
| **Travel Services**   | | | |
| expedia.com           | 52.30%    | 43.79%    | 19.4%        |
| orbitz.com            | 42.95%    | 39.08%    | 9.9%         |
| travelocity.com       | 43.04%    | 39.20%    | 10.3%        |

Table 8: Proportion of households who visit a site 48 hours prior to ultimately making a purchase. The values are are broken down into two groups: those where the purchase is made quickly within duration $D \leq 10$ minutes, and those made slowly in $D > 10$ minutes. The final column reports the percentage difference between these two proportions.
| Segment 1 | Segment 2 | Segment 3 |
|-----------|-----------|-----------|
| **Marginal Means** | | |
| $E(S)$ | $0.85$ | $3.88$ | $9.28$ |
| $Pr(B=1)$ | $0.0132$ | $0.0502$ | $0.0880$ |
| $E(P)$ | 8.832 pages | 16.490 pages | 29.108 pages |
| $E(D)$ | 5.796 mins | 10.213 mins | 15.886 mins |
| $E(D/P)$ | 0.8557 pgs/min | 0.8096 pgs/min | 0.7483 pgs/min |
| $E(S/P)$ | 0.07 $/page$ | 0.20 $/page$ | 0.26 $/page$ |
| $E[\pi_j|y]$ | 0.3652 | 0.4448 | 0.1900 |

| **Mean of 2007 Internet Activity Variables** | | |
| Total Online Spend | $705.36$ | $896.17$ | $1128.65$ |
| Total no. Transactions | 8.44 | 10.38 | 12.96 |
| No. Top 100 Site Sessions | 1778 | 1801 | 1832 |
| No. Households Allocated | 6238 | 4972 | 1165 |

Table 9: Profile of the segments in the fitted three segment mixture model for oldnavy.com. Top Section: expected values of sales, purchase incidence, page views and duration, as well as the search velocity and and sales per page viewed. Bottom Section: means of the three household internet activity variables for households allocated to each segment. Bottom Row: number of households in the calibration dataset allocated to each segment using the mode of the household specific posterior probabilities.
Figure 2: Expected sale amount and purchase probabilities per visit at amazon.com and bn.com for the stochastic model. Panel (a) plots expected sale amount conditional on visit duration; Panel (b) plots expected sale amount conditional on the number of page views; Panel (c) plots expected sale amount conditional on visit duration for situations where a purchase is made (i.e., $B = 1$); Panel (d) is the purchase probability against the visit duration; Panel (e) is the purchase probability against the number of page views. Ninety percent confidence intervals, calculated using the bootstrap, are plotted as light shaded intervals in each panel and for both websites.
Figure 3: Relationships between sales and visitation variables for online apparel retailers in panels (a), (c) and (e), and online travel service providers in panels (b), (d) and (f). Panels (a) and (b) present the expected spend conditional on duration of visit $E(S|D)$. Panels (c) and (d) present the expected spend conditional on duration and that a purchase is made, $E(S|B = 1, D)$. Panels (e) and (f) depict purchase probability conditional on the number of page views, $\Pr(B = 1|P)$. 
Figure 4: Estimated relationships for apple.com. Panels (a) and (b) plot the purchase probability against visit duration and the number of page views. Panels (c) and (d) plot the expected spend against duration and number of page views for visits which result in a purchase.
Supplementary Materials for Panagiotelis, Smith and Danaher

Part A

In this part of the supplementary material we report some additional summary statistics for the nine datasets employed in the paper.

| Website          | Mean  | Standard Deviation |
|------------------|-------|--------------------|
|                  | S     | D     | P     | B     | S     | D     | P     | B     |
| Amazon           | 4.38  | 11.6  | 13.6  | 0.08  | 36.2  | 16.8  | 18.1  | 0.27  |
| Barnes & Noble   | 2.79  | 10.7  | 13.9  | 0.08  | 15.1  | 15.6  | 18.2  | 0.27  |
| Apple            | 0.52  | 12.8  | 8.8   | 0.29  | 1.9   | 17.2  | 10.0  | 0.45  |
| JC Penney        | 6.70  | 11.1  | 26.3  | 0.06  | 41.7  | 15.5  | 33.9  | 0.25  |
| Victoria’s Secret| 8.87  | 12.9  | 29.0  | 0.08  | 39.0  | 17.0  | 35.9  | 0.27  |
| Old Navy         | 3.64  | 10.1  | 17.4  | 0.05  | 22.2  | 13.9  | 20.8  | 0.21  |
| Expedia          | 10.90 | 11.7  | 11.5  | 0.03  | 105.9 | 15.8  | 12.7  | 0.16  |
| Travelocity      | 12.97 | 11.1  | 12.3  | 0.03  | 116.2 | 14.9  | 14.4  | 0.17  |
| Orbitz           | 13.31 | 10.2  | 11.5  | 0.03  | 109.5 | 14.3  | 12.8  | 0.18  |

| Website          | Skew Coefficient | Maximum Value |
|------------------|------------------|---------------|
|                  | S     | D     | P     | S     | D     | P     |
| Amazon           | 54.0  | 3.4   | 3.5   | 6031.8| 179  | 489  |
| Barnes & Noble   | 20.2  | 3.4   | 3.6   | 1161.8| 179  | 199  |
| Apple            | 23.6  | 3.0   | 3.6   | 249.0 | 179  | 310  |
| JC Penney        | 22.4  | 3.4   | 3.4   | 3361.0| 179  | 407  |
| Victoria’s Secret| 7.5   | 3.1   | 3.2   | 1328.7| 177  | 472  |
| Old Navy         | 11.1  | 3.4   | 3.4   | 757.5 | 174  | 375  |
| Expedia          | 21.4  | 3.2   | 3.2   | 9262.9| 179  | 278  |
| Travelocity      | 21.5  | 3.2   | 3.4   | 8913.4| 175  | 259  |
| Orbitz           | 16.5  | 3.3   | 3.5   | 5604.0| 178  | 303  |
Part B

In this part of the supplementary material we first specify the likelihood of the mixture model used in Section 4, augmented with the latent variables \( L \). Then we outline a Bayesian MCMC sampling scheme to generate draws from the augmented posterior distribution. Last, we show how the output of the sampling scheme can be used to compute predictions of the sales variables, conditional on the visit variables. The approach differs slightly depending upon whether, or not, the household in the holdout sample also has visits recorded in the calibration dataset.

Let \( y_i = \{S_i, B_i, P_i, D_i\} \) be the \( i \)th observation of the sales and visitation variables, and assume that this observation comes from household \( h(i) \). We denote the data as \( y \), and the set of all stochastic model parameters, excluding \( \pi \) and \( \alpha \), as \( \Phi \). Then the augmented likelihood for \( n \) observations is

\[
L_A(\Phi, L; y) = \prod_{i=1}^{n} \prod_{l=1}^{M} f^l(y_i|\Phi^l)^{\mathcal{I}(L_h(i) = l)},
\]

where \( f^l(y_i|\Phi^l) \) is the density corresponding to the distribution function \( F^l \), \( \Phi^l \) are the parameters of mixture segment \( l \) only, and \( \mathcal{I}(A) = 1 \) if \( A \) is true, and zero otherwise. The contribution to the augmented likelihood of \( f^l \) is

\[
f^l(y_i|\Phi^l) = f_3^l(P_i)f_2^l(B_i|P_i \in \tilde{P}_k) \prod_{j=0,1} f_1^l(S_i, D_i|B_i = j, P_i \in \tilde{P}_k)^{\mathcal{I}(B_i = j)},
\]

where \( f_1^l, f_2^l, f_3^l \) are the density and probability mass functions corresponding to the distribution functions \( F_1^l, F_2^l, F_3^l \) as specified in the paper.

The following MCMC sampling scheme generates iterates repeatedly and sequentially from each of the steps below until convergence to the joint posterior distribution. After this a Monte Carlo sample of the parameters and latent variables is collected. We note that this sampling scheme applies where a Log-Logistic distribution is employed to model sales.

**Sampling Scheme for Latent Class Model**
(1) Generate from $L_h|\Phi, \pi, y$, for all households $h = 1, \ldots, H$

(2) Generate from $\pi|\Phi, L, \alpha, y$

(3) Generate from $\alpha|\Phi, L, \pi, y$

(4) Repeat for each mixture component $l = 1, \ldots, M$:

   (a) Generate $r, p$ for the modified NBD $\mathcal{F}_3^l$

   Repeat for each page view partition $k = 1, \ldots, K$:

   (b) Generate $p_B$ for the Bernoulli $\mathcal{F}_2$

   (c) Generate $\mu_0, \lambda_0$ for the Inverse Gaussian $\mathcal{F}_1|B = 0$

   (d) Generate $\mu_1, \lambda_1$ for the Inverse Gaussian $\mathcal{F}_1|B = 1$

   (e) Generate $\mu_s, \sigma_s$ for the Log-Logistic $\mathcal{F}_s$

   (f) Generate $\theta$ for the Gaussian copula $\mathcal{C}$

Steps 1-3 generate the latent variables and parameters associated with the mixture model from their Bayesian conditional posterior distributions. Step 4 generates the parameters of the mixture components. This is repeated for all $M$ mixture components, and also in Steps 4(b) to 4(f) for all $K$ page view partitions, although we avoid denoting the parameters with additional subscripts for clarity of exposition. Overall, there are $M(2+8K)$ parameters generated in Step 4.

Deriving the conditional posteriors in Step 4 involves standard Bayesian calculations. Where the posteriors cannot be recognised as known distributions we use the random walk Metropolis-Hastings algorithm to generate the parameters. This is a very widely used Bayesian tool, see Robert and Cassella (2004, pp. 287-291) for an introduction. We outline below how to undertake Steps 1 to 3 which are less standard. Let

$$L_h(l) = (L_1, \ldots, L_{h-1}, L_h = l, L_{h+1}, \ldots, L_H)$$

denote the latent variable vector with household $h$ allocated to class $l$, so that $L_h = l$. Then,
the conditional posterior in Step 1 is

\[
Pr(L_h = l | \Phi, \pi, y) = \frac{\mathcal{L}_A(\Phi, L_h(l); y)\pi_l}{\sum_{j=1}^M \mathcal{L}_A(\Phi, L_h(j); y)\pi_j}.
\]

In Step 2, the conditional posterior of the vector of class probabilities is \( \pi \sim \text{Dirichlet}(\tilde{\alpha}) \), where \( \tilde{\alpha} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_M) \), \( \tilde{\alpha}_l = \alpha_l + n_l(L) \) and \( n_l(L) \) is the number of observations in class \( l \) for the classification given by \( L \). In Step 3 \( \alpha \) is generated using the random walk Metropolis-Hastings algorithm.

Below we show how to evaluate the terms required to compute the forecasts in Section 4.2. The mass function

\[
f(B_i, \pi, \Phi, L_h = l | D_i, P_i, y) \propto f_{1D}(D_i | B_i, P_i; \Phi^l) f_2^l(B_i | P_i; \Phi^l) f_3^l(P_i; \Phi^l) f(\Phi, \pi, L_h = l | y)
\]

where \( f_{1D}, f_2^l \) and \( f_3^l \) are the known component density and mass functions for segment \( l \).

Computing the above when \( B_i = 1 \) and \( B_i = 0 \), and normalising, gives

\[
Pr(B_i = 1 | D_i, P_i, y) = \int \sum_{l=1}^M f(B_i = 1, \pi, \Phi, L_h = l | D_i, P_i, y) d(\Phi, \pi)
\]

\[
= \int \sum_{l=1}^M g(\Phi, L_h = l) f(\Phi, \pi, L_h = l | y) d(\Phi, \pi),
\]

where \( g(\Phi, L_h = l) = A'(B_i = 1; \Phi^l)/[A'(B_i = 1; \Phi^l) + A'(B_i = 0; \Phi^l)] \). Note that the summation and integral of \( g \) are weighted by the augmented posterior distribution \( f(\Phi, \pi, L_h = l | y) \).

This can be approximated as

\[
Pr(B_i = 1 | D_i, P_i, y) \approx \frac{1}{J} \sum_{j=1}^J g(\Phi_{ij}, L_{ij}^{[j]}),
\]

using the Monte Carlo iterates \{\( \Phi_{ij}, \pi_{ij}, L_{ij}^{[j]} \}_{j=1}^J \}. However, if no visits by household \( h \) are
recorded in the calibration data, then \( L_h \) is not generated as part of the sampling scheme discussed above. In this case, \( L_h^{[j]} \) is generated from the multinomial distribution with parameters \( \pi^{[j]} = (\pi_1^{[j]}, \ldots, \pi_M^{[j]}) \) from sweep \( j \) of the sampling scheme.

The expected sale amount of the visit is

\[
E(S_i|D_i, P_i, y) \approx \int \sum_{l=1}^M E(S_i|D_i, P_i, \Phi, L_h = l) f(\Phi, \pi, L_h = l|y) d(\Phi, \pi),
\]

which can be approximated using the Monte Carlo sample as

\[
E(S_i|D_i, P_i, y) \approx \frac{1}{J} \sum_{j=1}^J E(S_i|D_i, P_i, \Phi^{[j]}, L_h^{[j]}).
\]

The expectation inside the summation above can be computed using numerical integration as discussed in Section 2.4. However, to repeat this \( J \) times is computationally demanding. Therefore, we employ the histogram estimate \( E(S_i|D_i, P_i, y) \approx \frac{1}{J} \sum_{j=1}^J S_i^{[j]} \), where \( S_i^{[j]} \sim f(S_i|D_i, P_i, \Phi^{[j]}, L_h^{[j]}) \). Let \( L_h^{[j]} = l \), then to generate \( S_i^{[j]} \) we first draw \( B_i^{[j]} \sim f(B_i|D_i, P_i, \Phi^{[j]}, L_h^{[j]} = l) \), which is given Section 2.4, and then \( S_i^{[j]} \sim f(S_i|B_i^{[j]}, D_i, P_i, \Phi^{[j]}, L_h^{[j]} = l) \), which can be derived simply from the bivariate copula model. As before, if no visits by household \( h \) are recorded in the calibration data, then \( L_h^{[j]} \) is generated from the multinomial distribution with parameters \( \pi^{[j]} \) from sweep \( j \) of the sampling scheme.