ISSN 1440-771X

MONASH University

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

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

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

February 2013

Working Paper 05/13

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

Anastasios Panagiotelis
Department of Econometrics and Business Statistics,
Monash University, Australia.
Email: Anastasios.Panagiotelis@monash.edu

Michael S. Smith
Melbourne Business School,
The University of Melbourne, Australia.
Email: mike.smith@mbs.edu.

Peter J. Danaher
Department of Marketing,
Monash University, Australia.
Email: Peter.Danaher@monash.edu

26 February 2013

JEL classification: C51, M31

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

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

February 26, 2013

Anastasios Panagiotelis is Lecturer in Econometrics at Monash University, Michael S. Smith is Professor of Econometrics at Melbourne Business School 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 used for part of this study. 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 construct 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 parametric distributions that account for consumer heterogeneity, and involves copula components. Our model is readily estimated using full maximum likelihood, 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. 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 account for latent household segments, further accounting for consumer heterogeneity. The model can also be adjusted to accomodate 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 stochastic model, when compared to regression, probit regression and a popular data-mining method.

Keywords: Copulas, Heterogeneity, Marketing Models, Online Purchasing, Panel data.

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 for each these variables are very different, and there is no known suitable multivariate distribution to employ. Therefore, we construct one by composition, where the component distributions are drawn from the marketing and economic literatures, and also shown to fit well empirically. A key component is the bivariate distribution of sales and duration, conditional on page view 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

Embrechet 2005), they have only recently been employed in marketing models; see Danaher and Smith (2011) and Glady, Lemmens and Croux (2010) for examples. An advantage of the model is that it can be estimated rapidly using full 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 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 three main research questions.

The first research question is to 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 simultaneously determined and therefore endogenous. We also show how our stochastic model can be adjusted to account for online retailers with products at a very limited number price points, such as apple.com, where the primary sale item is a $0.99 song. To demonstrate the effectiveness of the stochastic model, we use it to forecast both purchase incidence and sales amount, given the website visit variables, in a validation study. For all nine websites examined the forecasts prove to be more accurate than those from a variety of alternative approaches, including a naïve benchmark, regression modeling and a popular data mining method.

Our second research question is to examine these relationships empirically, and investigate to what extent they are similar, or vary, 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 both complex and nonlinear throughout. It suggests that page views are a stronger determinant of sales and purchase incidence than visit duration, as suggested, but not verified, by Lin et al. (2010). 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 of these two websites as a function of visit duration and page views is nearly identical. However, there can be strong differences across product categories. For example, there is evidence that consumers research 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 decreases 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.

Our last research question is to assess and capture consumer heterogeneity. This is partially accounted for at the observation level by using a Negative Binomial Distribution (NBD) for page views, and an Inverse Gaussian distribution for duration, in the composition. These are distributions widely used for data that exhibit substantial heterogeneity. To account for further heteorgeneity at the household level, we extend our model to a latent class finite mixture model. The mixture components are quad-variate distributions of the type developed here by composition, and estimation is via Bayesian Markov chain Monte Carlo methodology. Using the example of oldnavy.com, we show that there is evidence of two distinct market segments: a large group of nonbuying browsers, and a smaller group of more goal-directed and higher-spending buyers that are more efficient in their navigation through the website.

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 its advantages over alternative approaches. We show how the proposed model can be estimated, both in the case where sales amounts are continuously distributed, as well as for the case where retailers, such as

Apple, offer products mostly at a few discrete price points. From the stochastic 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 show dominates competing forecasts based on regression modeling and data mining techniques.

2.1 ComScore Data

The data used in this study were collected by comScore and made available by subscription via the Wharton Research Data Service (WRDS). The database comprises a randomly-selected subsample of 100,000 members of a panel of over two million internet users from across the United States. These users were observed continuously over 2007, during which all website visitation and online transaction activity was captured passively using proprietary software that is installed on individual machines in a household at the time of recruitment. 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.[1] 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

[1] It is important to note that we do not have clickstream data. The page view and duration values are the sum of, repectively, the pages viewed and length of time on each URL within the domain name. For example, if a vistor goes to amazon.com and clicks though 5 pages and spends 10 seconds on each page, the recorded page views is 5 and the duration is 50 seconds.

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. The number of observations in our dataset is 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 total internet activity of the full ComScore panel, which contains millions of observations for each of the largest retailers.

2.2 Stochastic Model

We build a stochastic model for 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 $S \geq 0$. This requires constructing the joint distribution of all four variables, which is difficult because of their very different natures. 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 finanical (McNeil, Frey and Embrechts 2005) analysis, although they have only been employed recently in the marketing literature (Danaher and Smith 2011). Their most attractive characteristic is they permit the combination of any univariate marginal distributions that need not come from the same distributional family, yet still allow for dependence among variables. However, in this situation such a direct use of a copula model to account for dependence 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, popular elliptical copulas such as the Gaussian and t copula (McNeil, Frey and Embrechts 2005; p.191) only have 6 and 7 parameters, respectively, while most Archimedean copulas only have a single dependence parameter. Overall, standard copula models are insufficently parameterized to capture the dependence structure between these four variables, which we show is nuanced in our empirical work.

Instead, we construct the joint distribution via composition as

$$F(S, B, D, P) = F_1(S, D|B, P)F_2(B|P)F_3(P)\,, \qquad (2.1)$$

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, any nonlinearity or other complexities in the dependence between the sales and website visit variables can be uncovered. We derive the purchase probability and expected sales, 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, as we show below, there are parametric distributions for each component in equation (2.1) that have been widely-used to model similar variables previously, and they also fit well for our website data. 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 full 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, which has been used previously by Danaher (2007) and Huang and Lin (2006). Since this distribution can be derived as a Gamma-Poisson mixture, it implicitly accounts for 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 adjust the NBD probability mass function g to remove the zero case, resulting in a probability mass function of $\Pr(P = p) = g(p)/(1 - g(0))$, where $g(\cdot)$ is the probability mass function of the NBD.

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 $\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_K$ 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. In our empirical analysis we set $K = 10$, which is a fine enough grid to capture the variation in parameter values, but coarse enough to ensure that the sample sizes within our partitions are sufficiently large to estimate the distributional parameters. We show that this greatly enhances the quality of fit, as well as substantially improves prediction accuracy in a validation study. Conditional on the page view partition, we model the purchase indicator, $F_2(B|P \in \tilde{P}_k)$, as simply 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 distribution. This distribution is widely used to model duration over heterogeneous populations (Hougaard 1984; Johnson, Kotz and Balakrishnan 1994, p.291), which is precisely the situation here.[2]

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 view partition. We "couple" these together univariate distributions using a bivariate copula function C with dependence parameter θ_k,

[2] The Inverse Gaussian was also identified as the best fitting distribution to duration in our website data using AIC from a list of alternatives that included the Gamma, Weibull, Log-Logistic, Inverse Gaussian and Log-Normal distributions.

which differs for each page view partition $k = 1, 2, \ldots, K$. The copula model expresses the bivariate distribution as

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

There are many choices for the bivariate copula function that can be employed here, with comprehensive lists given by Nelsen (2006), McNeil et al. (2005) and Trivedi and Zimmer (2005). However, a particularly popular and versatile choice is the bivariate Gaussian copula, which is defined as

$$C(u, v; \theta) = \Phi_2(\Phi_1^{-1}(u), \Phi_1^{-1}(v); \theta)\,, \qquad (2.3)$$

where Φ_1^{-1} is an inverse standard normal distribution function and Φ_2 is the distribution function of a bivariate normal distribution with zero mean, unit marginal variances and correlation $-1 < \theta < 1$. It is important to make clear here that using a Gaussian copula function does not mean that (S, D) is distributed Gaussian. Instead the copula simply 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 an extensive discussion of the Gaussian copula function in two or more dimensions.

The resulting distribution F_1 can be shown to always have $F_{1S}(S|B = 1, P \in \tilde{P}_k)$ and $F_{1D}(D|B = 1, P \in \tilde{P}_k)$ as its two marginal distributions. For the univariate distribution of visit duration we again employ an Inverse Gaussian distribution, but for sales we employ a Log-Logistic distribution. The Log-Logistic has long been employed to model the distribution of income in economics, and has also been used successfully to model sales (Oyer 2000).[3]

For apple.com, sales occur at a very limited number of price points, with 87.33% of all purchases being for exactly \$0.99, which corresponds to the sale of a single song from the

[3] The Inverse Gaussian for duration, and the Log-Logistic for sales, were also identified using AIC as the optimal choices here for our website data from a list of alternatives that included the Gamma, Inverse Gaussian, Weibull, Log-Logistic and Log-Normal distributions.

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 the ordinal-valued distribution, with values at all the unique price points observed in the data. 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. Overall, 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. In our empirical work $K = 10$, so that we estimate 82 parameters from the data in our stochastic model.

2.3 Estimation

Another benefit of employing the decomposition in equation (2.1) is that estimation can be undertaken using maximum likelihood on each component distribution, and the resulting point estimates are the maximizers of 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 $B = 1$ the bivariate copula model in equation (2.2) is estimated as discussed in Cherubini, Luciano and Vecchiato (2004; pp.154-156). The ease with which bivariate copulas can be estimated, whether the Gaussian copula or another copula, is a further reason for their popularity.

In the case of discrete pricing, estimation of the stochastic model is unchanged, except for the bivariate Gaussian copula. As noted by Danaher and Smith (2011), maximum likelihood estimation for the Gaussian copula parameter θ is very different when one of the margins is discrete, which is the case here. 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 ith observation of the pair (S, D), $u_{D,i} = \hat{F}_{1D}(d_i|B = 1, P)$, $b_i = \hat{F}_{1S}(s_i|B = 1, P)$ and $a_i = \hat{F}_{1S}(s_i^-|B = 1, P)$ be the left hand limit of the step function $\hat{F}_{1S}$ at s_i. Then the likelihood for the bivariate copula is[4]

$$f(\theta; \text{data}) = \prod_i \left(\tilde{C}(b_i|u_{D,i}; \theta) - \tilde{C}(a_i|u_{D,i}; \theta)\right) f_{1D}(d_i|B = 1, P)\,,$$

where $\tilde{C}(u_1|u_2; \theta) = \frac{\partial}{\partial u_2} C(u_1, u_2; \theta)$ and the product is taken only over observations where sales are made ($B = 1$) and with page views in the partition $P \in \tilde{P}_k$. The likelihood is easily maximized with respect to $-1 < \theta < 1$ by simple grid search.

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 (Nelsen 2006, Chapter 6), which is exactly the case for our data. Computation of Spearman's rho for the distribution at equation (2.1) can be undertaken in a Monte Carlo fashion by first simulating many iterates from this distribution, and then computing the correlation coefficient of the ranked iterates.[5]

The pairwise dependence between S and D can also be computed from the bivariate copula model. Using the copula parameter θ_k, for a bivariate Gaussian copula model, Spear-

[4] For the distribution function $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$, where X_1 is discrete valued and X_2 continuous, the density is obtained by differentiation with respect to x_2 and differencing with respect to x_1. This gives mixed probability density $f(x_1, x_2) = (\tilde{C}(b|u_2) - \tilde{C}(a|u_2))f_2(x_2)$, where $\tilde{C}(u_1|u_2) = \frac{\partial}{\partial u_2} C(u_1, u_2)$, $b = F_1(x_1)$ and $a = F_1(x_1^-)$, which is the left hand limit of F_1 at x_1.

[5] To simulate an iterate from $F(S, B, D, P)$ 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).

man's rho is given by $\rho_k^C = (6/\pi)\text{arcsin}(\theta_k/2)$ (Cherubini et al. 2004, p. 104). This is the measure of dependence of the distribution $F_1(S, D|B = 1, P \in \tilde{P}_k)$, and is therefore conditional on a purchase being made and the number of page views being in the kth range in the partition. We label it with a superscript "C" to distinguish it from the marginal Spearman's rho.

The impact of page views and duration can be understood in greater detail by evaluating the expected sales amount and probability of purchase, conditional on the visit variables. These can be computed from the stochastic model as follows. Using Bayes rule, the probability of a purchase can be expressed as

$$\Pr(B = 1|D, P) = \frac{f_{1D}(D|B = 1, P)\Pr(B = 1|P)}{f_{1D}(D|B = 1, P)\Pr(B = 1|P) + f_{1D}(D|B = 0, P)\Pr(B = 0|P)} . \quad (2.4)$$

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 $E(S|D, P) = \int sf(s|D, P)\text{d}s$ is obtained via univariate numerical integration, except in the case of apple.com where summation over the discrete domain of sales can be used instead. The density function $f(s|D, P)$ 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 obtained from equation (2.4). The remaining density can be computed from the copula model [6] as

$$\begin{aligned} f(s|B = 1, D, P) &= \frac{f_1(s, D|B = 1, P)}{f_{1D}(D|B = 1, P)} \\ &= c\left(F_{1S}(s|B = 1, P), F_{1D}(D|B = 1, P); \theta\right) f_{1S}(s|B = 1, P) , \end{aligned}$$

[6] Note that a copula model with continuous margins and bivariate distribution $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ has density $f(x_1, x_2) = c(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2)$, with $c(u_1, u_2) = \partial^2 C(u_1, u_2)/\partial u_1 \partial u_2$ and marginal densities $f_1(x_1) = \text{d}F_1(x_1)/\text{d}x_1$ and $f_2(x_2) = \text{d}F_2(x_2)/\text{d}x_2$.

where f_1, f_{1D} and f_{1S} are the density functions of distributions F_1, F_{1D} and F_{1S}, and $c(u, v; \theta)$ is the so called "copula density" (Cherubini et al. 2004; pp.81-84). For the bivariate Gaussian copula this is given by

$$c(u, v; \theta) = \frac{\partial^2 C}{\partial u \partial v} = (1 - \theta^2)^{-1/2} \exp\left\{\frac{-\theta^2(w_u^2 + w_v^2) + 2\theta w_u w_v}{2(1 - \theta^2)}\right\} ,$$

where $w_u = \Phi^{-1}(u)$ and $w_v = \Phi^{-1}(v)$.

We note that standard estimation methods for regression models to compute $E(S|D, P)$, and binary regression models for $\Pr(B|D, P)$, do not account for any endogeneity and can produce biased estimates. An advantage of computing the conditional expectation $E(S|D, P)$ and probability $\Pr(B = 1|D, P)$ from the stochastic model at equation (2.1) is that it captures directly the contemporaneous 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 choice of stochastic model, as discussed below.

2.5 Model Validation

We demonstrate that our stochastic model improves prediction of the two sales variables, conditional on both visit variables, compared to a number of alternative approaches. These include a naïve forecast as a benchmark, regression modeling, a popular data-mining method, and a stochastic model where the parameters of F_1 and F_2 are constant with respect to page views.

For seven of the nine websites that we examine in detail we select a holdout sample of $n_f = 5,000$ observations. As amazon.com and apple.com have the largest number of observations, we consider a larger holdout sample of $n_f = 10,000$ for these two websites. The holdout sample is stratified with respect to purchase incidence (B) and page view partition (P), so that it is more representative than a simple random sample.[7] We fit our stochastic model to the data for each website, excluding the holdout sample. From the fitted model we compute the probability of a purchase, $\hat{b}_i = \Pr(B = 1|D_i, P_i)$, and expected spend, $\hat{s}_i = E(S|D_i, P_i)$, 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, and the root mean square errors are calculated for the holdout samples for each of the nine websites as:

$$\text{RMSE}(B) = \left(\frac{1}{n_f} \sum_{i=1}^{n_f} (B_i - \hat{b}_i)^2 \right)^{1/2} , \quad \text{RMSE}(S) = \left(\frac{1}{n_f} \sum_{i=1}^{n_f} (S_i - \hat{s}_i)^2 \right)^{1/2} .$$

We label our method 'SM2', and compare it to five alternative methods which construct forecasts as follows:

- *Naïve*: The historical purchase incidence ($\hat{b}_i = \bar{B}$) and the average sales amount of transactions multiplied by the historical purchase incidence $\hat{s}_i = (\bar{S}|B = 1) \times \hat{b}_i$.

- *Regression*: 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 covariates. We then compute the probability of a sale $\hat{b}_i = \Pr(B = 1|D_i, P_i)$ from the probit model, and expected spend 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 spend for a transaction with given duration and page views from the regression model.

- *CART1*: We employ the popular "Classification and Regression Tree" (CART) data-mining method with D and P as input variables; once for purchase incidence, and a second time for sales amount, including visits that do not conclude in a transaction.

- *CART2*: We employ the same CART model for purchase incidence as above, but then fit CART a second time to the logarithm of spend of transactions only. Forecasts for sales amount for all visits are then computed by multiplication in the same manner as

[7] The conclusions are unaffected by the manner in which the holdout samples are selected at random.

for the regression models.

- *SM1*: This is our stochastic model, but where the parameters of F_1 and F_2 are assumed constant with respect to page views P (i.e., P is not partitioned into deciles).

Table 3 summarises the predictive accuracy of all six methods over the holdout samples for all nine websites. It is clear that duration of a visit and the number of page views are useful in forecasting sales incidence and amount, with the naïve forecasts dominated by at least one method that uses D and P as inputs. For every website the proposed stochastic model (SM2) outperforms the alternative methods. Interestingly, SM2 outperforms SM1 throughout, which shows the necessity of making the parameters of F_1 and F_2 functions of the conditioning variable P in Section 2.

3 Empirical Analysis

We now use our stochastic model to investigate 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 examine the case of apple.com separately, given its unique online retail product assortment and 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 (comprising 47% of total 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 4 contains the page view partitions, and the number of observations in each page view range.

Table 5 reports the parameter estimates and confidence intervals for the component distributions in equation (2.1), and it is easily seen that all the parameters vary significantly over the page view ranges.[8] 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 10 is low at $\Pr(B = 1|1 \leq P \leq 10) = 1.27\%$. As one would expect, this increases monotonically with the number of page views, so that for visits with more than 67 page views the probability of a purchase is much higher, at $\Pr(B = 1|P \geq 67) = 30.39\%$. For each bivariate Gaussian copula we also compute Spearman's rho for the dependence between sale amount and visit duration, assuming a purchase does occur. This is also reported in Table 5 and is positive for all page view partition ranges, with the lowest value being $\hat{\rho}_8^C = 0.035$ and the largest being $\hat{\rho}_1^C = 0.079$.

Thus, it might initially appear that dependence between the duration of a visit and the sale amount is quite low. However, Table 6 reports the matrix of pairwise marginal Spearman's rhos, and the marginal Spearman dependence between duration and sale amount is $\hat{\rho}_{S,D} = 0.2592$. For comparison, Table 6 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 amount spent (S) at each visit.

We also compute the expected spend, 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 spend for each page view level.

Figure 1 shows that the link between duration and expected sales is very different to that between page views and expected sales. We examine this further by computing the expected

[8] Estimation took 26 seconds in Matlab, highlighting the computational viability of the approach for the large datasets that arise in the study of online retail.

spend, marginalizing, respectively, over page views and duration, as detailed in Appendix A.[9] Starting with the expected sale amount and duration relationship (marginalizing over page views), denoted $E[S|D]$, Figure 2(a) shows that expected sales grow rapidly to \$14.68 for visits of duration of 77 minutes, then plateaus. The high Spearman's marginal pairwise correlation of $\rho_{P,D} = 0.6247$ between visit duration and page views is therefore caused by visits of durations less than 77 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 in the stochastic model as outlined in Appendix A. 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, and while other segments that 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 the stochastic model can be extended to incorporate latent segmentation household-level heterogeneity.

For comparison we also look at barnesandnoble.com, which is the website of Barnes and

[9] These expectations can be computed in closed form using the stochastic model adopted.

Noble, the second largest online book retailer in the U.S. 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 4 shows that the partitions for the page view deciles for these two retailers are broadly similar.

Figure 2 also plots the expected spend and probability of purchase for visits to barnesandnoble.com, conditional on the website visit variables for the fitted model. In comparison to amazon.com expected spend peaks for slightly shorter visits of duration 69.6 minutes, but at a much lower value of just over $10.23. 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 spend does not increase as quickly with duration as for amazon.com. However, there appears to be little difference in the way the two websites produce a purchase. 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 spend appears to result from higher basket totals for sales at amazon.com. Interestingly, visits with very similar durations of 41.2 and 43.6 minutes have the maximum purchase probability for amazon.com (0.22) and barnesandnoble.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 spend 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 spend. The site travelocity.com is most successful, with the highest expected spend of $143.09 for visits of duration 121 minutes. The differences between the three sites appear driven entirely by differing abilities to convert longer duration visits into purchase events. Once the expected spend 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 reflects the homogeneity of their products.

Figures 3(e) and 3(f) depict the probability of a sale being made against the number of page views for the six websites. Apart from very high page view values for expedia.com, higher page views correspond to a higher probability of purchase throughout. Interestingly, the sites that are least successful at converting longer duration visits into sales, are not necessarily poor at converting page views into more purchases. For example, the probability of a purchase during a visit to oldnavy.com is the highest of all retailers as the number of page views increases. It seems reasonable that the number of page views is more closely related to the marginal costs of website delivery than the visit duration, so that by this measure oldnavy.com is the most efficient of the three apparel retailers.

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 plausable 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 7 reports these proportions, and a clear pattern emerges for the different product categories. Fast purchasers of apparel products research online 48 hours 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), supporting 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.

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 retailers, but these convert much more rapidly into higher purchase probabilities than other retailers in our study. The expected spend 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 spend against duration and page views, respectively. The expected spend 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)). For all these other retailers, visits where a purchase is made has higher expected spend as duration increases. For visits with a small number of pages views, the expected spend at apple.com is close to $0.99. However, this increases markedly for visits with 10 or more page views.

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 6. This is particularly true for the purchase indicator, with $\hat{\rho}_{B,D} = 0.217$ and $\hat{\rho}_{B,P} = 0.174$. 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

While the NBD and Inverse Gaussian distributions account for observation-level heterogeneity, to account for further household-level consumer heterogeneity we consider 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. Finite mixture models are often estimated using the EM algorithm (McLachlan and Peel 2000), although Bayesian estimation using Markov chain Monte Carlo (MCMC) has gained in popularity because it allows computation of the full range of posterior inference (Diebolt and Robert 1994; Richardson and Green 1997). This includes the ability to profile the segments, which can be difficult using other likelihood-based methods (Wedel and DeSarbo 2002).

4.1 Bayesian Finite Mixture Model

Consider a mixture model with M latent segments, where the probability that a household is a member of segment l is π_l. Then, for household h, the joint distribution of the website visit and sales variables is

$$G_h(S,B,D,P) = \sum_{l=1}^{M} \pi_l F^l(S,B,D,P) = \sum_{l=1}^{M} \pi_l \left(F_1^l(S,D|B,P) F_2^l(B|P) F_3^l(P) \right) . \qquad (4.1)$$

This is a mixture of M stochastic models, each of the type defined in equation (2.1) and with mixture component (i.e., segment) membership denoted with a superscript.

To estimate the mixture model latent multinomial variables are introduced to specify segment membership for each household, 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, all observations are allocated to one of the M segments, which makes it much easier to estimate the parameters of each component F^l. Following Diebolt and Robert (1994), Lenk and DeSarbo (2000) and others we use a MCMC algorithm that explicitly generates the latent variables, and then the parameters of each mixture component conditional on segment membership. A similar approach has also proven popular in Bayesian estimation of choice models (Albert and Chib 1993; Edwards and Allenby 2003). [10]

Prior Distributions

To define a Bayesian model the prior distributions of all the parameters in the model have to be specified, along with those of any hyperpriors. We adopt a Dirichlet prior for $\pi = (\pi_1, \ldots, \pi_M) \sim \text{Dirichlet}(\alpha)$, which is the most common choice in mixture modeling because it has the Bayesian property of being conjugate to the multinomial (Diebolt and Robert 1994). To make the mixture model more flexible, we make α a hyperparameter with a uniform hyperprior on $[0,2]^M$, which ensures the prior on π is flat at the prior expected value of

[10] For a general introduction to Bayesian MCMC estimation and inference, including its particular suitability for models involving latent variables such as that here, see Robert and Casella (2004) and Gamerman and Lopes (2006).

$E(\alpha) = 1$. 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.

The Fitted Mixture Model

Appendix B outlines a MCMC sampling scheme to generate K iterates from the posterior distribution of the mixture model parameters, augmented with the latent variables. When employing this we were careful to check that there was no evidence of "label switching" in the output of the sampling scheme. Label switching is a well-known potential pitfall in Bayesian estimation of finite mixture models; see Lenk and DeSarbo (2000) and Stephens (2000) for an outline and discussion of the problem. Using the iterates, Bayesian posterior inference can be computed in a Monte Carlo fashion. This includes parameter estimates, but of particular interest in this study are the profiles of the market segments. One advantage of Bayesian estimation is that 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 y denotes the data and Φ^l the parameters of the lth segment, then the mean of the lth fitted segment is

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

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

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

$$\begin{aligned} \Pr(L_h = l|y) &= \int \Pr(L_h = l|\pi, \Phi, y) f(\pi, \Phi|y) \mathrm{d}(\pi, \Phi) \\ &\approx \frac{1}{K} \sum_{k=1}^{K} \Pr(L_h = l|\pi^{[k]}, \Phi^{[k]}, y) = \hat{\omega}_h^l \,. \end{aligned} \tag{4.2}$$

Here, $\Phi = \{\Phi^1, \ldots, \Phi^M\}$, $\{\pi^{[k]}, \Phi^{[k]}\}$ are the Monte Carlo iterates output from the MCMC scheme, and Appendix B outlines how to compute the probability in the summation 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 π_l in equation (4.1), which is not household specific. Last, we identify the number of components M in the mixture model using BIC, which is likely to correspond to using the exact model posterior probabilities as advocated by Lenk and DeSarbo (2000) because of the large sample sizes used here.

4.2 Segmentation for OldNavy.com

To demonstrate, we fit mixture models with up to four segments for oldnavy.com. The two segment model had the lowest BIC value, and Table 8 gives profiles of these two segments.[11] The top portion of the table reports the marginal expectations of our four key variables and the two ratios P/D and S/P. The first ratio is a measure of how fast the 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. Members of the first segment have an expected spend of \$7.21 per visit, make purchases 8.2% of the time, visit on average for 14.22 minutes and navigate through 25.6 pages. In comparison, on average, members of the second segment spend much less (\$1.59), make substantially fewer purchases (2.6%), visit for shorter periods (7.97 minutes) and navigate fewer pages (12.5). They account for 70% of visitors, probably indicating surfing rather than buying behavior for these households. Members of the first segment also appear more goal-directed than those in the second segment, with

[11] The BIC values for the one to four segment models are 827321, 824494, 825411 and 826306, respectively.

higher expected spend-per-page (\$0.248 compared to \$0.124), but with comparable search velocity (2.82 compared to 2.93 pages per minute). Figure 5 plots the expected spend and probability of purchase against visit duration for both segments. This supports the idea that households in the first segment are the serious customers, with purchase incidence growing much faster with duration of visit.

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 8 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. On average, Segment 1 households out-spend those from Segment 2, but only by a factor of about 1/4; whereas it is by a factor of around 4 at oldnavy.com. Nevertheless, overall internet activity is comparable between the two segments (1761 and 1788 sessions), so that households in the first segment appear more goal-directed online customers in general.

5 Conclusion

In this research we develop a stochastic model for website visit duration, pages viewed, 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 of the two "stickiness" measures, page views is a better indicator 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. 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 underlying purchase probability is the same for book and apparel websites; something that is not readily apparent from a naïve analysis of sales alone. For example, the difference in sales amounts across apparel websites is more likely due to the product offering, rather than characteristics of people visiting the websites. Our latent class segmentation for oldnavy.com reveals two distinct market segments. The larger segment consists of low spend visitors who exhibit browsing behaviour, while the smaller segment consists of more engaged customers who exhibit greater goal-directed behavior. A study of the wider online activities of these visitors 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 components that are drawn from the marketing and economic literatures, and fit the online data 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 numerical integration. Third, computing these expectations from the quad-variate distribution, rather than modeling them directly in a regression style framework, accounts for the endogeneity in the simultaneous determination of all four variables. Fourth, consumer heterogeneity is accounted for at the visit level by the use of the NBD and Inverse Gaussian distributions, and at the household level by the extension to a finite mixture of the quad-variate distributions. Last, we adapt the stochastic model to cope with the discrete pricing used by retailers such as apple.com. To acheive this we exploit the flexibility of the bivariate copula component to model a mixture of continuous and discrete marginals; something that would otherwise be difficult. To futher extend this to the household-level latent segmentation model in Section 4 only requires a minor adjustment of the Bayesian approach, where a data augmentation method can be used to account for the discrete margin as outlined in Smith and Khaled (2012).

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 stochastic model, and that the two website stickiness measures provide valuable information when predicting purchase incidence and sales. Future work in this area 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 conceptionally straightforward by making the parameters functions of the covariates. In addition, because we have a stochastic model, a possible further extension is a hierarchical model that incorporates household-level heterogeneity. This would provide an alternative to latent class segmentation.

Appendix A: 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 visitation variable and marginalized over the second. Marginalizing over the number of page views can be achieved as follows:

$$\begin{aligned}
\Pr(B=1|D) &= \sum_{P=1,2,\ldots} \Pr(P, B=1|D) \\
&= \sum_{P=1,2,\ldots} \frac{f(P, B=1, D)}{f(D)} \\
&= \frac{\sum_{P=1,2,\ldots} f(P, B=1, D)}{\sum_{P=1,2,\ldots} \sum_{B=0,1} f(P, B, D)} \\
&= \frac{\sum_{P=1,2,\ldots} f(D|B=1, P)\Pr(B=1|P)\Pr(P)}{\sum_{P=1,2,\ldots} \sum_{B=0,1} f(D|B, P)\Pr(B|P)\Pr(P)} \,.
\end{aligned}$$

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

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

Each component in this summation is known from the stochastic model definition and the expression computed analytically. Derivation of the expected spend conditional on duration only is similar. We first note that $E(S|D) = \sum_{B=0,1} \sum_{P=1,2,\ldots} E(S|D, P, B)\Pr(P, B|D)$, but because $S = 0$ when $B = 0$, this expression simplifies to

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

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

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

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

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

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

Appendix B: Mixture Model Estimation

In this appendix we first specify the likelihood of the mixture model 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.

Let $y_i = \{S_i, B_i, P_i, D_i\}$ be the ith 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 π and α, as Φ. Then the augmented likelihood for n observations is

$$\mathcal{L}(\Phi, L; y) = \prod_{i=1}^{n} \prod_{l=1}^{M} f^l(y_i|\Phi^l)^{\mathcal{I}(L_{h(i)}=l)} ,$$

where $f^l(y_i|\Phi^l)$ is the density corresponding to the distribution function F^l in equation (4.1), Φ^l are the parameters of mixture component 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 in equation (4.1) and are specified in Table 2. We note that the likelihood of the stochastic model in Section 2 is equivalent to that of a single component with $M = 1$.

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 but that a similar algorithm could be developed for the discrete pricing case.

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 F_3^l

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

        (b) Generate p_B for the Bernoulli F_2

        (c) Generate μ_0, λ_0 for the Inverse Gaussian $F_{1D}|B = 0$

        (d) Generate μ_1, λ_1 for the Inverse Gaussian $F_{1D}|B = 1$

        (e) Generate μ_s, σ_s for the Log-Logistic F_{1S}

        (f) Generate θ for the Gaussian copula 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

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

In Step 2, the conditional posterior of the vector of class probabilities is $\pi \sim$ 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 α is generated using the random walk Metropolis-Hastings algorithm.

References

Albert, James and Siddhartha Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data", *Journal of the American Statistical Association*, 88, 669-679.

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

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.

Edwards, Y. and Greg Allenby (2003), "Multivariate Analysis of Multiple Response Data", *Journal of Marketing Research*, 40, 321-334.

Gamerman, Dani and Hedibert Freitas Lopes (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, (2nd ed.), CRC Press.

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.

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

McLachlan, Geoffrey and David Peel (2000), *Finite Mixture Models*, Wiley: NY.

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.

Richardson, Sylvia and Peter J. Green (1997), “On Bayesian Analysis of Mixtures with an Unknown Number of Components”, *Journal of the Royal Statistical Society*, Series B, 59, 4, 731-792.

Robert, Christian and George Casella (2004) *Monte Carlo Statistical Methods*, (2nd ed.), New York, NY: Springer.

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

Smith, Michael S. and Mohamad Khaled (2012), "Estimation of Copula Models with Discrete Margins via Bayesian Data Augmentation", *Journal of the American Statistical Association*, forthcoming.

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

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.

	Ranking by Total Sales		Ranking by Total Number of Purchases	
Rank	Domain Name	Sales ($1000)	Domain Name	Purchases
1	amazon.com	1921.1	apple.com	12.65%
2	southwest.com	1838.1	amazon.com	7.61%
3	expedia.com	1682.7	ups.com	4.13%
4	dell.com	1623.8	walmart.com	2.77%
5	orbitz.com	1188.0	jcpenney.com	2.37%
6	ticketmaster.com	1151.9	victoriassecret.com	2.02%
7	travelocity.com	1138.6	oldnavy.com	1.92%
8	jcpenney.com	831.0	staples.com	1.60%
9	cheaptickets.com	801.5	safeway.com	1.50%
10	walmart.com	751.7	yahoo.net	1.50%
11	aa.com	718.2	intuit.com	1.46%
12	staples.com	638.0	officedepot.com	1.46%
13	delta.com	632.8	quillcorp.com	1.42%
14	continental.com	623.4	orientaltrading.com	1.41%
15	qvc.com	607.5	papajohnsonline.com	1.38%
16	victoriassecret.com	598.9	yahoo.com	1.35%
17	quillcorp.com	547.4	qvc.com	1.16%
18	yahoo.com	543.2	vistaprint.com	1.15%
19	jetblueairways.com	500.3	columbiahouse.com	1.07%
20	apple.com	497.1	quixtar.com	1.02%

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.

Component	Distribution	Parameters	Density/Probability Mass	Domain
$F_3(P)$	Modified NBD	r, q	$\Pr(P=p)=\frac{g(p)}{1-g(0)}$, with $g(p)=\begin{pmatrix} r+p-1 \\ p \end{pmatrix} q^r(1-q)^p$	$p=0,1,\ldots$
$F_2(B\mid P)$	Bernoulli	p_B	$f_2(b)=p_B^b(1-p_B)^{(1-b)}$	$b\in\{0,1\}$
$F_{1D}(D\mid B=0,P)$	Inverse Gaussian	μ_0,λ_0	$f_{1D}(d)=\sqrt{\frac{\lambda_0}{2\pi d^3}}\exp\left\{-\frac{\lambda_0}{2\mu_0^2 d}(d-\mu_0)^2\right\}$	$d>0$
$F_{1D}(D\mid B=1,P)$	Inverse Gaussian	μ_1,λ_1	$f_{1D}(d)=\sqrt{\frac{\lambda_1}{2\pi d^3}}\exp\left\{-\frac{\lambda_1}{2\mu_1^2 d}(d-\mu_1)^2\right\}$	$d>0$
$F_{1S}(S\mid B=1,P)$	Log-logistic	μ_S,σ_S	$f_{1S}(s)=\frac{\exp((\log(s)-\mu_S)/\sigma_S)}{s\sigma_S(1+\exp((\log(s)-\mu_S)/\sigma_S))^2}$	$s>0$
C	Gaussian Copula	θ	$c(u,v;\theta)=(1-\theta^2)^{-1/2}\exp\left\{\frac{-\theta^2(w_u^2+w_v^2)-2\theta w_u w_v}{2(1-\theta^2)}\right\}$	$(u,v)\in[0,1]^2$

Table 2: Component distributions in the stochastic model for the joint distribution of (S,B,D,P) in Section 2. Apart from the modified NBD, the parameters vary over the $K=10$ page view partitions. For the Gaussian copula $w_u=\Phi^{-1}(u)$ and $w_v=\Phi^{-1}(v)$, with Φ the standard normal distribution function. Note that the Log-Logisitic distribution is replaced by the Empirical Distribution Function in the case of apple.com to account for highly discrete pricing.

	Purchase Incidence (RMSE(B))						Spend (RMSE(S))					
Retailer	Naïve	Regression	CART1	CART2	SM1	SM2	Naïve	Regression	CART1	CART2	SM1	SM2
Books & Digital Media												
barnesandnoble.com	0.266	0.257	0.270	0.270	0.255	**0.246**	14.40	13.98	16.06	14.30	14.01	**13.78**
amazon.com	0.268	0.261	0.257	0.257	0.258	**0.248**	23.36	23.31	25.42	26.76	22.92	**22.57**
apple.com	0.453	0.444	0.442	0.442	0.442	**0.439**	1.544	1.539	1.543	1.495	1.536	**1.464**
Travel Services												
expedia.com	0.159	0.156	0.160	0.160	0.154	**0.149**	97.63	97.19	103.7	99.23	95.19	**93.59**
travelocity.com	0.174	0.171	0.178	0.178	0.167	**0.164**	104.9	108.1	110.1	106.0	101.9	**100.7**
orbitz.com	0.179	0.171	0.184	0.184	0.170	**0.167**	114.8	109.2	120.2	115.4	110.4	**108.9**
Apparel												
victoriassecret.com	0.272	0.258	0.275	0.275	0.256	**0.252**	39.40	38.46	39.28	38.35	37.36	**37.05**
oldnavy.com	0.215	0.202	0.222	0.222	0.203	**0.198**	18.93	18.58	20.98	20.12	18.10	**17.86**
jcpenney.com	0.245	0.236	0.250	0.250	0.234	**0.232**	40.29	39.81	43.75	41.14	39.30	**39.14**

Table 3: Predictive performance of the different methods for all nine websites in the validation study. The numbers in the table are the root mean square errors of forecasts for purchase incidence (RMSE(B)), and also sale amount (RMSE(S)). Smaller values correspond to more accurate forecasts in the holdout sample, and the best performing method in each case is in bold.

		amazon.com			barnesandnoble.com	
k	$\tilde{P}_k$	No Sale	Purchases	$\tilde{P}_k$	No Sale	Purchases
1	1-10	257004	3306	1-12	25440	321
2	11-14	35785	4054	13-16	2700	356
3	15-17	17009	2964	17-19	1315	276
4	18-20	12481	2640	20-22	1060	265
5	21-24	11828	3050	23-26	985	268
6	25-29	10164	3148	27-31	831	296
7	30-36	9035	3323	32-38	794	283
8	37-46	7881	3053	39-49	708	282
9	47-66	7484	3133	50-70	593	290
10	67-500	7283	3180	71-500	567	286
Total		375954	31851		34993	2923

Table 4: Page view partitions for the two book retailers. Also given are the sample sizes for each partition, broken down by observations where no sale was made, and those where one or more items were purchased.

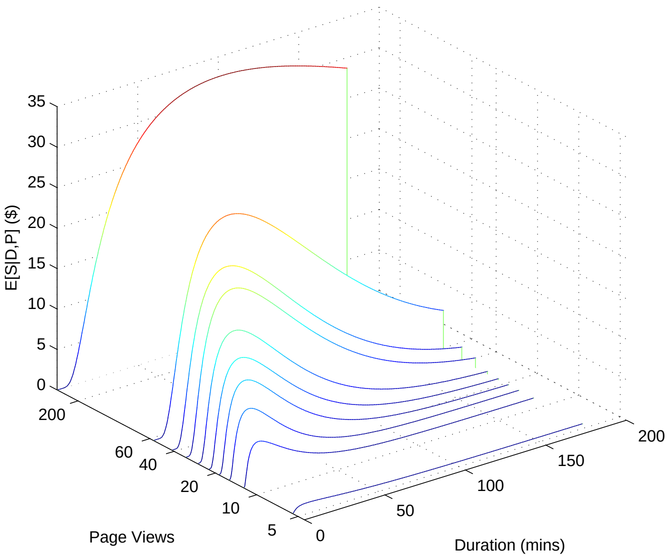

Figure 1: Expected sale amount at amazon.com as a function of duration of visit and number of page views (on the logarithmic scale) resulting from the parametric stochastic model.

Page Views	Copula Dependence		$F_{1S}(S \mid B=1,P)$		$F_{1D}(D \mid B=1,P)$		$F_{1D}(D \mid B=0,P)$		$F_2(B \mid P)$
(P)	$\hat{\theta}$	$\hat{\rho}^C$	$\hat{\mu}_S$	$\hat{\sigma}_S$	$\hat{\mu}_1$	$\hat{\lambda}_1$	$\hat{\mu}_0$	$\hat{\lambda}_0$	$\hat{p}_B$
$1 \leq P \leq 10$	0.083	0.079	3.220	0.472	6.26	7.25	5.15	3.19	0.0127
	[0.049 0.117]		[3.192 3.248]	[0.459 0.486]	[6.07 6.46]	[6.90 7.60]	[5.13 5.18]	[3.18 3.21]	[0.012 0.013]
$11 \leq P \leq 14$	0.045	0.043	3.272	0.486	10.13	18.41	11.79	11.69	0.1018
	[0.014 0.075]		[3.247 3.298]	[0.474 0.499]	[9.90 10.36]	[17.61 19.21]	[11.66 11.91]	[11.52 11.86]	[0.099 0.105]
$15 \leq P \leq 17$	0.073	0.070	3.316	0.492	12.80	29.03	14.95	17.31	0.1484
	[0.037 0.109]		[3.285 3.347]	[0.477 0.507]	[12.50 13.11]	[27.55 30.50]	[14.74 15.16]	[16.95 17.68]	[0.144 0.153]
$18 \leq P \leq 20$	0.069	0.066	3.410	0.499	15.69	38.25	16.90	21.45	0.1746
	[0.031 0.107]		[3.378 3.443]	[0.484 0.516]	[15.31 16.08]	[36.19 40.32]	[16.64 17.17]	[20.91 21.98]	[0.169 0.181]
$21 \leq P \leq 24$	0.077	0.074	3.438	0.497	18.42	48.25	19.79	27.85	0.2050
	[0.042 0.112]		[3.408 3.469]	[0.48 0.512]	[18.01 18.82]	[45.83 50.67]	[19.49 20.09]	[27.14 28.56]	[0.199 0.212]
$25 \leq P \leq 29$	0.044	0.042	3.481	0.502	21.86	63.26	22.70	36.09	0.2365
	[0.009 0.079]		[3.451 3.512]	[0.488 0.517]	[21.41 22.30]	[60.13 66.38]	[22.35 23.04]	[35.09 37.08]	[0.229 0.244]
$30 \leq P \leq 36$	0.043	0.041	3.573	0.509	26.72	80.84	26.71	44.65	0.2689
	[0.009 0.077]		[3.543 3.603]	[0.495 0.524]	[26.20 27.24]	[76.95 84.73]	[26.24 27.14]	[43.35 45.95]	[0.261 0.277]
$37 \leq P \leq 46$	0.037	0.035	3.619	0.508	31.73	108.5	31.88	58.89	0.2792
	[0.001 0.072]		[3.588 3.650]	[0.493 0.523]	[31.12 32.33]	[103.1 114.0]	[31.36 32.40]	[57.05 60.73]	[0.271 0.288]
$47 \leq P \leq 66$	0.039	0.038	3.669	0.543	41.78	157.5	40.45	80.33	0.2951
	[0.004 0.074]		[3.636 3.702]	[0.527 0.559]	[41.03 42.53]	[149.7 165.4]	[39.80 41.10]	[77.76 82.90]	[0.286 0.304]
$67 \leq P \leq 500$	0.075	0.072	3.824	0.585	66.32	222.2	61.28	140.90	0.3039
	[0.040 0.109]		[3.788 3.859]	[0.569 0.603]	[65.01 67.58]	[211.3 233.1]	[60.35 62.20]	[136.3 145.5]	[0.295 0.313]
Parameters of the NBD F_3: $\hat{r} = 0.7660\,(\pm 4.27 \times 10^{-6})$; $\hat{q} = 0.0600\,(\pm 2.97 \times 10^{-7})$; $\widehat{E(P)} = 13.56$; Std. $\widehat{\text{Dev.}}(P) = 15.033$									

Table 5: Estimates of the parameters of the stochastic model for amazon.com, with 95% confidence intervals given below in parentheses.

	S	B	D	P
	Spearman Correlations			
S	1	0.9985	0.2592	0.3130
B		1	0.2606	0.3118
D			1	0.6247
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 6: Marginal pairwise Spearman dependence measures from the fitted stochastic model for amazon.com, and the Pearson sample correlations.

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%
jcpenney.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 7: 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
	Marginal Means	
E(S)	\$7.21	\$1.59
Pr(B=1)	0.082	0.026
E(P)	25.63 pages	12.51 pages
E(D)	14.22 mins	7.97 mins
E(P/D)	2.82 pgs/min	2.93 pgs/min
E(S/P)	0.248 \$/page	0.124 \$/page
$E[\pi_j \vert y]$	0.298	0.702
	Mean of 2007 Internet Activity Variables	
Total Online Spend	\$1094.74	\$755.09
Total no. Transactions	12.46	8.95
No. Top 100 Site Sessions	1761	1788
No. Households Allocated	2201	10675

Table 8: Profile of the segments in the fitted two segment mixture model for oldnavy.com. Top Section: expected values of sales (dollars), purchase incidence, page views (number of pages) and duration (minutes), 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 allocated to each segment using the household specific posterior probabilities based on a cutoff of 0.5.

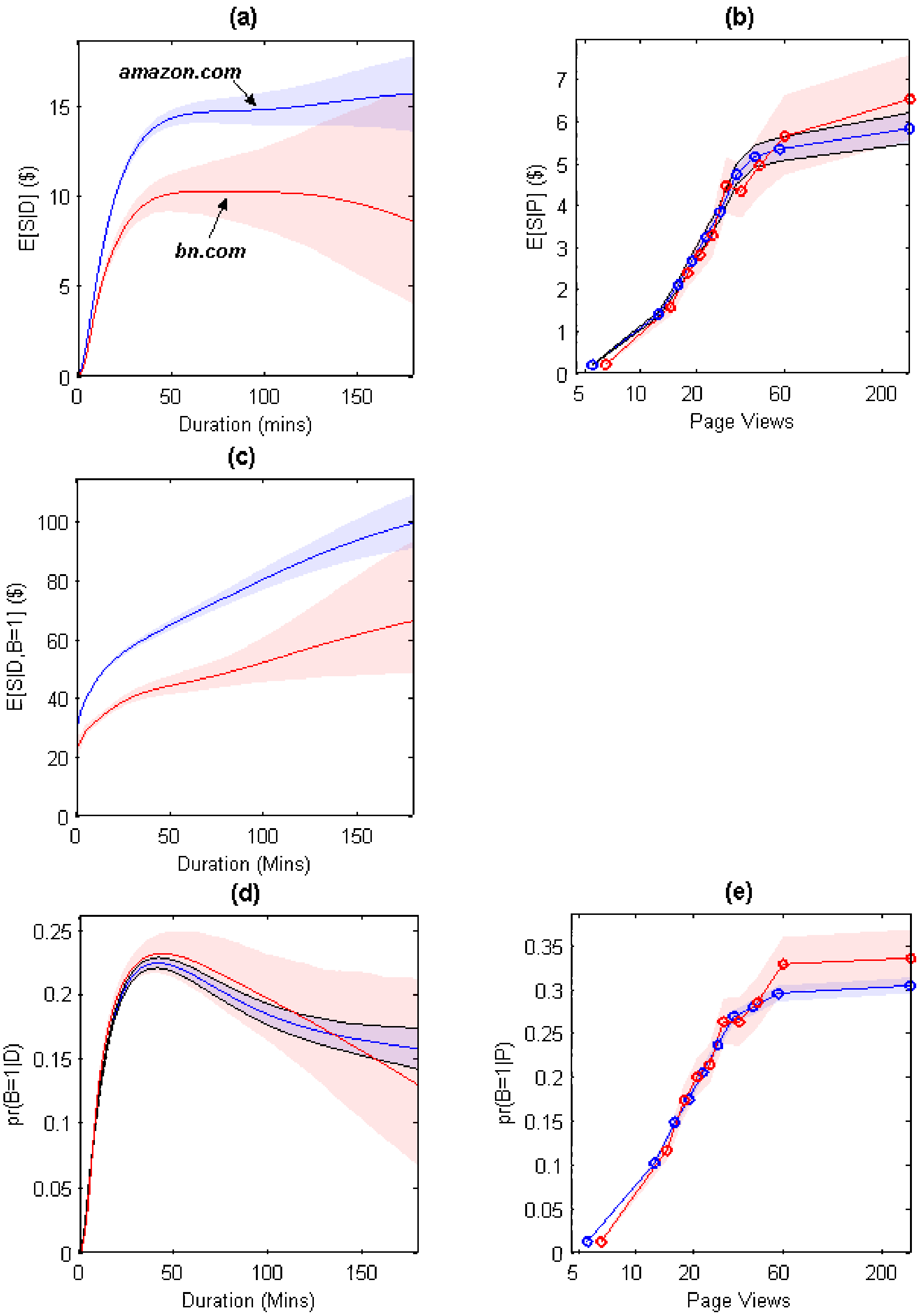

Figure 2: Expected sale amount and purchase probabilities per visit at amazon.com (blue line) and barnesandnoble.com (red line) for the stochastic model. Panel (a) plots expected sale amount conditional on visit duration; Panel (b) plots expected sale amount conditional on the number of page views; Panel (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.

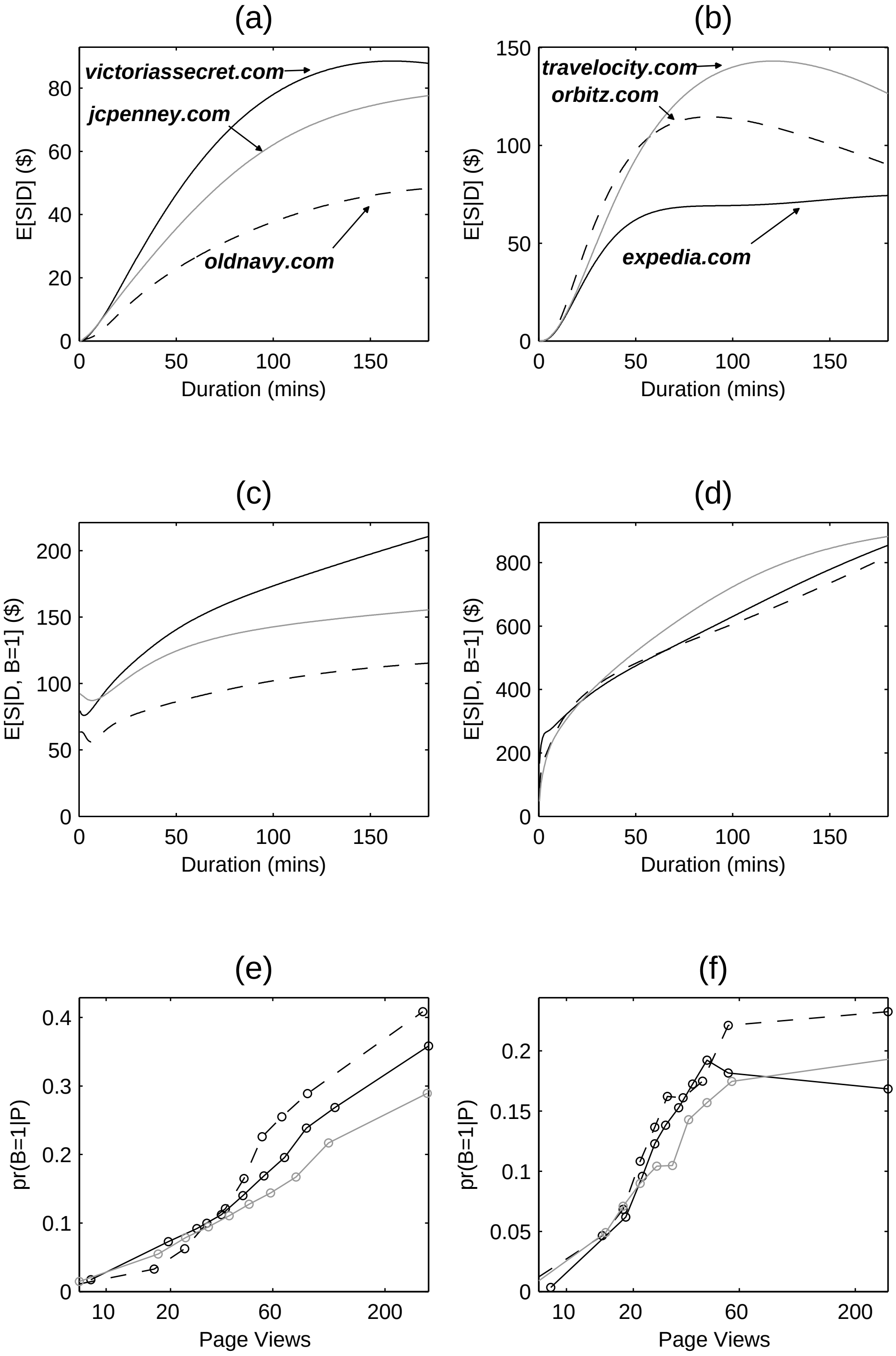

Figure 3: Relationships between sales and visitation variables for online apparel retailers in panels (a), (c) and (e), and online travel service providers in panels (b), (d) and (f). Panels (a) and (b) present the expected spend conditional on duration of visit $E(S|D)$. Panels (c) and (d) present the 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, $\Pr(B = 1|P)$.

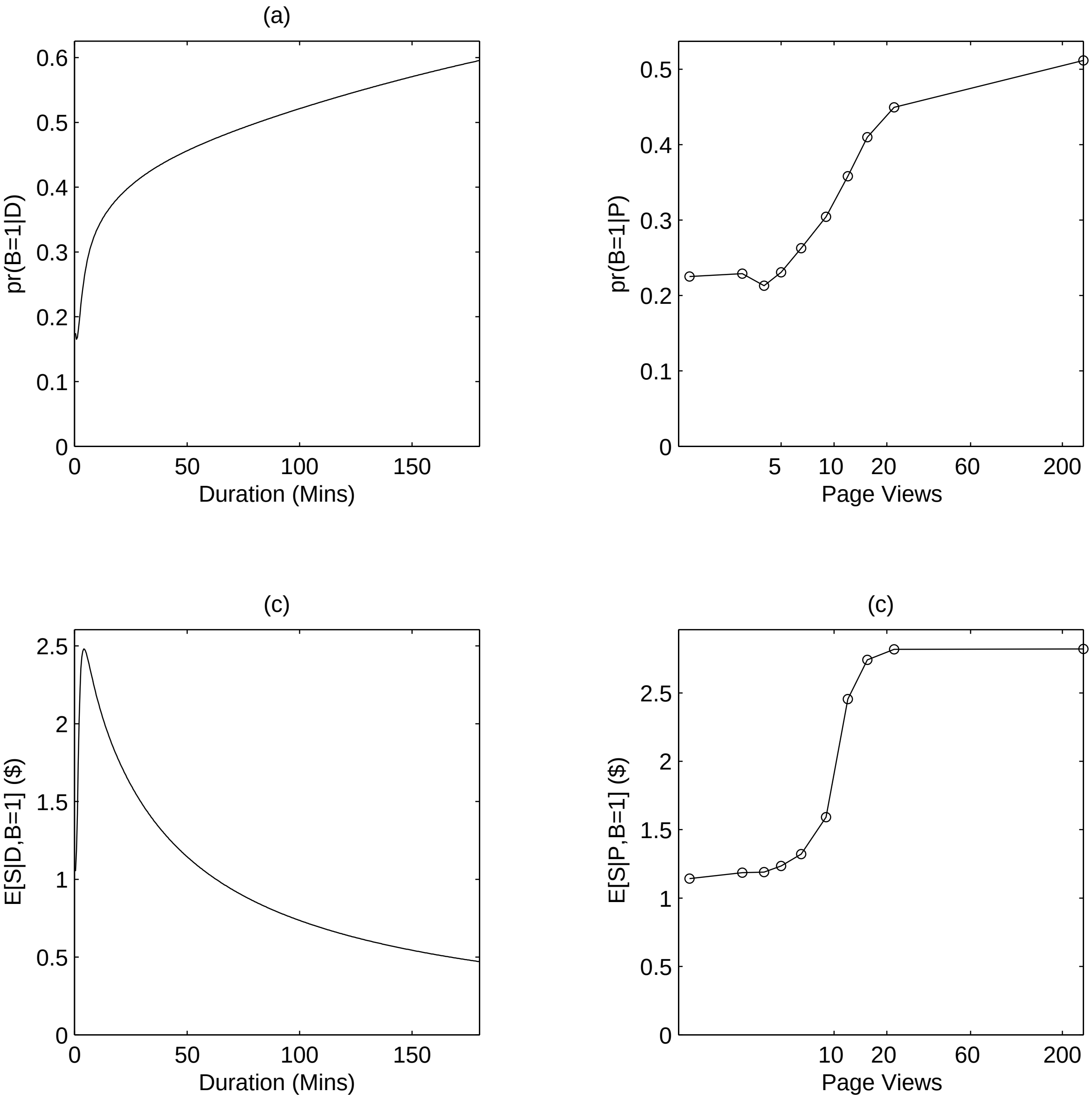

Figure 4: Estimated relationships for apple.com. Panels (a) and (b) plot the purchase probability against visit duration and the number of page views. Panels (c) and (d) plot the expected spend against duration and number of page views for visits which result in a purchase.

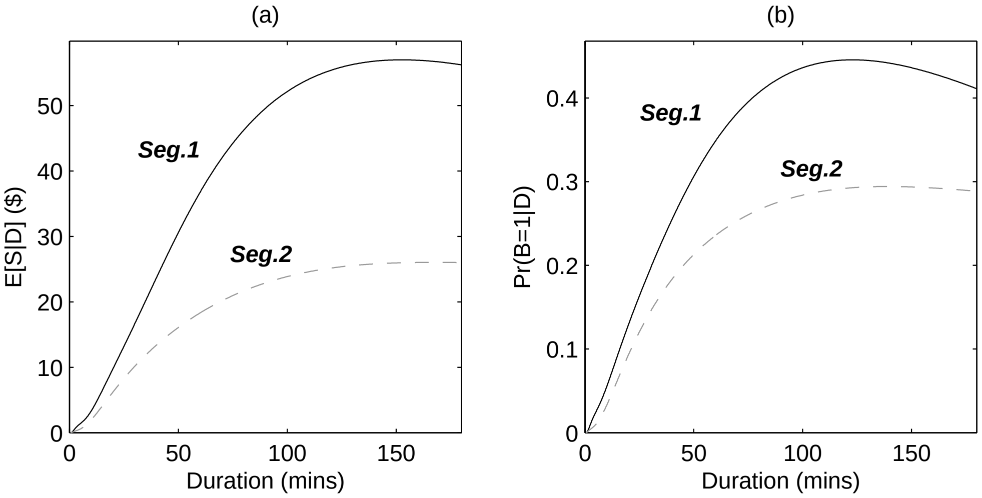

Figure 5: Relationship between sales and duration for oldnavy.com for the fitted two segment mixture model. Panel (a) gives the expected spend against duration, while panel (b) gives the probability of a purchase against duration. Each line corresponds to the relationship for a segment.