Multivariate Demand: Modeling and Estimation from Censored Sales

Catalina Stefanescu[*]

Abstract

Demand modeling and forecasting is important for inventory management, retail assortment and revenue management applications. Current practice focuses on univariate demand forecasting, where models are built separately for each product. However, in many industries there is empirical evidence of correlated product demand. In addition, demand is usually observed in several periods during a selling horizon, and it may be truncated due to inventory constraints so that in practice only censored sales data are recorded. Ignoring the inter-product demand correlation or the serial correlation of demand from one selling period to the next leads to biased and inefficient estimates of the true demand distributions. In this paper we propose a class of models for multi-product multiperiod aggregate demand forecasting. We develop an approach for estimating the parameters of the demand models from censored sales data in a maximum likelihood framework using the Expectation-Maximization (EM) algorithm. Through a simulation study, we show that the algorithm is computationally attractive and leads to maximum likelihood estimates with good properties, under different demand and censoring scenarios. We exemplify the methodology with the analysis of two booking data sets from the entertainment and the airline industries, and show that the use of these models in a revenue management setting for airlines increases the revenue by up to 11% relative to the use of alternative demand forecasting methods.

Key words: Demand estimation; multivariate models; maximum likelihood; EM algorithm; revenue management; retailing; inventory management.

[*] Management Science and Operations, London Business School, London NW14SA, United Kingdom. Email: cstefanescu@london.edu.

1 Introduction

Modeling and forecasting of customer demand is crucial in many application areas, including revenue management, retail assortment, and inventory management. In revenue management systems, demand forecasts are needed as inputs to any price optimization module (van Ryzin 2005), and sources from the airline industry estimate that a 20% reduction in demand forecast error may translate into a 1% increase in revenues (Talluri and van Ryzin 2003); in competitive industries with thin margins, such as airlines, a 1% improvement in revenues can be the difference between a successful and an unsustainable business. In retailing, knowledge of the true demand distribution and substitution rates is important for a wide range of category management decisions, such as the ideal assortment to carry, the optimal inventory to be stocked from each item, and the stock replenishment rates. Inventory management systems rely on accurate methods for estimating customer demand (Agrawal and Smith 1996), and their efficiency is particularly important for products (such as basic merchandise) where retailers have had increased competitive pressure.

There are two major challenges in modeling and forecasting customer demand. The first issue is that in practice different demand streams are often correlated, and demand models must account for this correlation. Patterns of demand correlation occur along two dimensions — the *time dimension* and the *product dimension*. Both types of demand correlation have been empirically documented in many industries, and they often arise in the same context.

The time dimension of demand correlation occurs when a firm sells products during a time horizon over which demand for a product recorded in different periods is related. For example, this is the case of travel tickets (e.g., train or airlines) sold during a booking period; tickets for travel during holiday times will be in higher demand throughout the booking horizon. In retailing, where cyclical demand fluctuations due to promotions occur commonly, inventory management decisions are often made periodically, hence there is a need for multi-period models that account for the serial correlation of demand from one period to the next.

The product dimension of demand correlation occurs when demand for different but related products is dependent. In magazine retailing, for example, Koschat (2008) finds evidence that demand for different magazines is correlated, and that a change in inventory levels of one magazine affects sales of the others. Correlation of product demands sometimes arises due to customer behavior. In retailing, correlated demand for color or style varieties of trendy apparel is a result of trend-following behavior. In general, when the retailer offers different styles, colors or flavors of the same product, substitution by the customer is likely to lead to correlated demands. Other instances of substitution which may induce demand correlation include buy-up (e.g., buying a higher fare ticket when the lower fares are not available) and buy-down (e.g., buying a lower fare ticket instead of a higher fare when the seller offers discounts). Note, however, that product demand correlation may happen not just due to substitution effects and other features of customer behavior, but also because products share some common characteristics. For example, demand for airline tickets on the London – New York route is likely to be high both in economy class and in business class, not primarily because of customer substitution (the high price difference will usually preclude this) but mainly since the two cities are both tourist and business destinations.

The second major issue in modeling and forecasting customer demand is estimating the parameters of demand models from censored sales data. In practice, only the recorded product sales are often available for estimation. However, actual demand may be greater than observed sales when a product sells out, hence sales data are just censored rather than exact observations of demand. Unobservable lost sales are prevalent in retailing where unmet demand arises when products are out-of-stock, particularly for low-cost, nondurable merchandise. If customers encounter a stockout for the product they desire, they may substitute with another product, place an order for delivery in the future, or move on without recording their request.[1] In the latter case, the sales data for the desired product are censored observations of demand.

Demand predictions based on sales data without accounting for the stockout effect potentially lead to two types of error. First, the forecasts for stocked-out products are negatively biased and the extent of the bias depends on the stockout incidence frequency. Wecker (1978) shows that stockouts also affect the estimate of the forecast error variance, and that the amount and direction of the effect depend on the stockout frequency, the coefficient of variation and the serial correlation of demand. In particular, he finds that the effect of stockouts on prediction accuracy is larger when demand has intertemporal correlation than when demand is uncorrelated between purchasing periods. The second type of error due to stockouts arises when customers purchase an alternative product and hence sales of substitute products increase. In this case, the estimates of ancillary demand for substitute products are positively biased.[2] Biased forecasts for the true demand based on censored sales data lead to a systematic decrease over time in the firm's expected revenue. This iteratively decreasing revenue pattern is similar to the spiral down effect investigated by Cooper et al. (2006) and

[1] Exceptions are catalogue ordering where the customer may place an order for a listed item that has meanwhile run out-of-stock, and e-retailing where the retailer does not reveal availability of a certain product before receiving a customer order. In this case the retailer can record the lost demand due to stockouts.

[2] As discussed earlier, product substitution by customers due to stockouts is also one potential source for correlation among observed sales levels.

caused by incorrect customer behavior assumptions inherent in many revenue management systems.

Demand estimation and forecasting from censored sales data must also be viewed in light of the price and demand relationship. When a product is not available, its price has essentially gone up to infinity. Moreover, even if the product is always available, price changes such as promotions usually have a substantial impact on sales levels. If price changes are not carefully tracked in the forecasting models, the uncensored demand estimates will suffer from fluctuations that can be at least partially explained by price changes. This effect is even more pronounced when considering substitute products and inter-temporal prices offered over a long sales horizon. It is therefore crucial to develop unconstraining techniques for estimating the parameters of *multivariate* demand models from censored sales data.

This paper addresses these issues and makes three contributions. First, we propose a class of multivariate demand models that capture both the time dimension and the product dimension of demand correlation. The models use the multivariate normal distribution to account for product demand correlation, and include a latent random term common to all time periods that induces the intertemporal demand correlation over the selling horizon. We discuss the patterns of correlation that can be captured with this class of models and show that they have the flexibility to cover a range of practical examples.

Second, we develop a methodology for estimating the parameters of demand models from censored sales data. Estimation is performed in a maximum likelihood framework using the Expectation-Maximization (EM) algorithm, first outlined by Dempster, Laird and Rubin (1977). In practice, convergence of the EM algorithm can be slow, particularly with large numbers of parameters or high degrees of censoring. We conduct a simulation study to investigate the properties of the EM estimates under different demand and censoring scenarios, and find that the algorithm converged within a reasonable running time in virtually all instances.

Third, we illustrate the methodology with applications to two industries, entertainment and airlines. For our first example, we use the modeling approach for analyzing a booking data set for performances at a London theatre. With relatively light censoring, we focus on ticket bookings in four price bands over seven periods, and we find that there is significant intertemporal demand correlation in each price band. We also document evidence of correlation of same period demand for tickets in different price bands, likely due to substitution effects. For our second example, we use the EM algorithm to estimate the parameters of demand models using airline booking data for two fare classes, over a booking horizon where many demand observations are censored due to lack of capacity. We find evidence of significant demand correlation for the two fare classes and across all booking periods. As a consequence of the high censoring incidence, the expected untruncated demand predicted by the model for all booking periods is much larger (up to 360%) than the estimates based on average censored sales. We also compare the untruncated demand predictions obtained with our methodology and with alternative models that ignore inter-temporal or inter-product correlation, and we find that in general our methodology leads to higher values of untruncated demand. Finally, we show how the multivariate demand models can be used to set protection levels for revenue management. Through a simulation experiment inspired by the airline booking data, we show that our demand modeling methodology leads in this setting to revenues up to 11% higher than those obtained using protection levels based on demand models that ignore intertemporal and inter-product correlation. In the highly competitive environment of the airline industry, such an improvement in revenue may be critical to the success of the company.

This paper is related to several different strands of literature. In biostatistics, reliability and economics, extensive research has focused on the estimation of distribution parameters from censored and truncated data — for good reviews see, for example, Lawless (2003) and Klein and Moeschberger (2005). The Kaplan-Meier estimator (Kaplan and Meier 1958) is the standard *nonparametric* procedure for estimating the distribution function of randomly censored univariate data. The method is statistically efficient and computationally simple, however it does not have a natural extension to the multivariate case. Moreover, nonparametric estimation methods have the general disadvantages that they cannot easily account for covariate effects, and they provide no basis for estimating the distribution beyond the censoring point (the stockout level). This is a major drawback for inventory management applications, since inventory stocking criteria rely on the tail of the demand distribution. On the other hand, *parametric* models can be estimated from censored data using hazard rate techniques in a lifetime framework, but most of these approaches have been developed for univariate data and it is difficult to extend them to multivariate distributions.

In the inventory management literature, Tan and Karabati (2004) provide a review on the estimation of demand distributions with unobservable lost sales. In particular, Nahmias (1994) assumes a model where demand follows a sequence of independent normal random variables, and examines three estimators for the mean and standard deviation of the demand distribution. Agrawal and Smith (1996) develop a parameter estimation method with lost sales when demand follows a negative binomial distribution, and show that this method is attractive for use in inventory replenishment applications. Lau and Lau (1996) discuss a procedure for estimating a univariate demand distribution from unobservable lost sales.

Lariviere and Porteus (1999) consider the case of one product with independent demand in different time periods following a newsvendor distribution, and discuss Bayesian updating of the demand model parameters at the beginning of each time period based on the observed sales during the previous period. All these papers assume univariate demand models and ignore correlation of product demand. A related stream of literature accounts for product demand correlation, while still ignoring time dependence. McGill (1995) considers a multivariate setting where single-period demand for different products is dependent. Anupindi, Dada and Gupta (1998) develop a model of choice between products that allows for substitution and lost sales in the event of a stock-out. They use an EM algorithm for estimating the demand parameters by treating the stock-out times as missing data, and find that demand rates estimated naively by using observed sales rates are biased even for items that have very few occurrences of stock-outs. Finally, Conlon and Mortimer (2007) estimate customer choice model parameters using the EM algorithm when no-purchase outcomes are unobservable.

In the revenue management literature most of the academic research has so far focused on pricing, assuming that the demand model is known. A few exceptions are van Ryzin and McGill (2000) who use the Kaplan-Meier method for unconstraining univariate demand, and Talluri and van Ryzin (2004) and Vulcano, van Ryzin and Ratliff (2008) who estimate customer choice demand models from sales data in the presence of stock-outs using the EM algorithm. Ratliff et al. (2008) overview the univariate demand untruncation literature in revenue management, with a focus on airline applications. Finally, Queenan et al. (2007) provide a review of unconstraining methods for the univariate demand models that have been used in revenue management practice.

This paper is also related to the literature on retail assortment planning with substitutable products. In this context, van Ryzin and Mahajan (1999) study a single-period assortment planning problem with a multinomial logit demand model allowing for assortment based substitution (when consumers substitute if the preferred product is not offered) but not for stockout-based substitution (when customers substitute when the preferred product is offered but temporarily unavailable), while Cachon, Terwiesch and Xu (2005) extend the model of van Ryzin and Mahajan (1999) to account for consumer search. These papers, however, do not consider the issue of modelling intertemporal demand correlation, and do not address the problem of estimating the parameters of the demand models from censored sales data. A recent paper by Kök and Fisher (2007) focuses on a periodic review inventory model with lost sales, develops an estimation approach for substitution rates from observed sales, and uses it to solve an assortment planning problem.

Most of the papers that focus on substitutable products in the context of retail assortment or revenue management, account for demand correlation between different products (but not between different time periods) by using customer choice models. These models reflect the way in which individual customers make their purchasing decisions, and have lately been the focus of increased research efforts. Their practical implementation, however, requires two different kinds of data for model calibration. First, the data must at least record the alternatives available to each customer at the time of the purchase request, as well as the final customer choice. This shopping alternatives data is often not available at the required level of detail; in particular, a capacity provider may not be able to record or even to observe all the alternatives offered to the customer when some of these alternatives are owned by competitors. Second, the customer population is usually heterogeneous and this heterogeneity has implications for customer choice demand modeling, as has long been documented in the marketing literature (Rossi, Allenby and McCulloch 2006). In such cases it is also necessary to account for the heterogeneity with good quality data on customer characteristics such as, for example, demographic variables, purchase history, or even geographical location. However, this customer specific data is also not always observable or recorded.

In summary, when the customer level and shopping alternatives data necessary for the calibration of customer choice models is easily available, these models are useful as they have great flexibility and forecasting power. When the required data is not available, however, multivariate models of *aggregate* demand are very useful as they can still capture both inter-product and intertemporal dependence patterns. This is the methodology that we investigate in this paper.

The remaining of the paper is structured as follows: Section 2 discusses the class of multivariate demand models and develops the estimation methodology. Section 3 presents the results of a simulation study and Section 4 shows the application of the methodology to the analysis of two booking data sets from the entertainment and airline industries. Section 5 concludes the paper with a discussion.

2 Model Specification and Estimation

In this section we first propose a class of multivariate demand models and discuss the patterns of correlation that they can capture. Next, we develop a maximum likelihood estimation methodology for the demand model parameters from censored sales data.

2.1 Model Specification

We consider the setting of a firm which sells n products over a time horizon $[0, T]$. Demand for each product is recorded at discrete time points t over the period $t = 1, \ldots, T$. Note that the selling periods do not need to be of the same length.[3] At any given time a product may be available for purchase or not, depending on the available inventory and on the product definition. For example, airlines usually open bookings for a flight up to one year in advance of the flight date, but certain fare classes have time of purchase restrictions and are no longer available close to departure. Let $D_t = (D_{t,1}, \ldots, D_{t,n})'$ denote the random vector of demand in period t, where $D_{t,i}$ is the demand for product $i \in \{1, \ldots, n\}$.

Let X_t be a $p \times n$ matrix of variables that influence product demand and that are directly observable, and let $\beta_t = (\beta_{t,1}, \ldots, \beta_{t,p})' \in \Re^p$ denote the vector of covariate effects parameters in period t. Each row of the X_t matrix corresponds to one of p covariates, and each column corresponds to one of the n products. These p covariates may include, for example, indicator variables for product restrictions, prices, as well as other product attributes and characteristics. The components of X may also be (nonlinear) functions of the covariates, rather than just the covariates themselves — for example, the logarithm of price usually gives a better fit to a linear demand model than the price itself. The components of X may be time-varying (for example, product restrictions may change during the selling horizon), or constant over time (product attributes such as color or style do not change from one period to the next). Typically, the first row of X_t will always contain ones, indicating the presence of an intercept term. Note that the values of some of the covariates are generally controlled by the product provider (for example, prices), while the values of other covariates may not be under the provider's control.[4]

We consider the following linear mixed effects model for the random demand in any period:

$$D_t = X_t'\beta_t + W_t \cdot v + \varepsilon_t, \quad t = 1, \ldots T. \tag{1}$$

The mean demand in period t is a linear function $X_t'\beta_t$ of known attributes. The random effects are modelled through an unobserved (latent) common shock $v \in \Re^n$ which

[3] Indeed in practice the selling periods often have variable lengths. Airline flights, for example, usually open for booking a year before the departure date. Flight demand forecasting then often uses booking information recorded during 24 periods over the one-year horizon, where a selling period can be as long as six months (at the beginning of the booking horizon), and as short as a day (close to the departure date).

[4] In a more general setting, the analysis may consider n products with the aim to forecast demand only for a subset of m products. For example, this is the case when the provider offers the m products which are related to the remaining $n - m$ products offered by competitors. In this case all the covariates would be observable, but the provider controls only the covariate values related to the m products that he offers.

influences demand in all periods. We assume that v has a multivariate normal distribution $v \sim \mathrm{N}(\mathbf{0},\ \mathbf{\Sigma}_v)$ with covariance matrix $\mathbf{\Sigma}_v$ and zero mean. The influence of the common random shock on demand for each product in period t is weighted by the $n \times n$ symmetric matrix W_t. The weighting matrices W_t are determined by the product definitions and are known by the modelers.[5] The error terms ε_t are normally distributed $\varepsilon_t \sim \mathrm{N}(\mathbf{0},\ \mathbf{\Sigma}_e)$ where $\mathbf{\Sigma}_e$ is a $n \times n$ diagonal matrix of error variances, so that in fact the components of the error vectors are independent. We also assume that ε_t are independent across time periods $t = 1, \ldots, T$, and independent of the random shock v.

With this specification, the latent random shock v induces correlation of demand both across different periods and across different products within the same period. Indeed, conditionally on v, the demand D_t has the multivariate normal $\mathrm{N}(X_t'\beta_t + W_t v,\ \mathbf{\Sigma}_e)$ distribution. Unconditionally, D_t has the $D_t \sim N(X_t'\beta_t,\ W_t'\mathbf{\Sigma}_v W_t + \mathbf{\Sigma}_e)$ distribution, hence demands at time t for different products are correlated. Also, for any distinct time periods $t \neq s$ we have

$$\mathrm{Cov}(D_t, D_s) = \mathrm{Cov}(W_t v,\ W_s v) = W_t \cdot \mathrm{E}[vv'] \cdot W_s = W_t \mathbf{\Sigma}_v W_s. \tag{2}$$

Thus the demand vectors D_t and D_s for different time periods are correlated because they share the influence of the common latent random shock v. To derive expression (2), recall that W_t is symmetric hence $W_t = W_t'$ for all t, and note that $\mathrm{E}[vv'] = \mathrm{Cov}(v, v) = \mathbf{\Sigma}_v$ since $\mathrm{E}[v] = 0$. From expression (2) it follows that higher diagonal values of $\mathbf{\Sigma}_v$ lead to stronger serial correlation of demand for any single product.

Model (1) can therefore account for both the time dimension and the product dimension of demand correlation. Different specifications of the structure of the covariance matrix $\mathbf{\Sigma}_v$ lead to a range of demand correlation patterns. In particular, the case of equal correlation of product demand can be modeled through the choice of an equicorrelated $\mathbf{\Sigma}_v$ matrix. The special case of independent product demand in all periods is obtained in model (1) when the covariance matrix of the random shock $\mathbf{\Sigma}_v$ is diagonal. The special case when single product demand is independent in different periods over the selling horizon is obtained when the corresponding diagonal component of matrix $\mathbf{\Sigma}_v$ is zero. Indeed, if $(\mathbf{\Sigma}_v)_{ii} = 0$, then product i demand in different periods does not share a random unobserved component and thus it has no serial dependence.

[5] For example, in airline bookings where products are defined as itinerary and fare class combination, the random shock will affect the demand for economy fare classes to a larger extent at the beginning of the booking period than at the end. Conversely, the shock will have a larger influence on the business demand closer to the time of service, rather than at the start of the booking period.

For notational convenience, we state the model for the demand over all time periods as

$$D = X'\beta + \mathbf{W}v + \varepsilon, \tag{3}$$

where $D, \varepsilon \in \Re^{nT}$ and $\beta \in \Re^{pT}$ are obtained by concatenating the corresponding vectors from periods $t = 1, \dots, T$, $X \in \Re^{pt \times nT}$ is the block matrix with matrices $X_1, \dots, X_T$ on the main diagonal and zeros elsewhere, and $\mathbf{W}$ is the $nT \times n$ matrix with rows given by $W_1, \dots, W_T$. Note that $\varepsilon \sim \mathrm{N}_{nT}(0, I_T \otimes \mathbf{\Sigma}_e)$, where $\otimes$ is the Kronecker product. We thus have $D \sim \mathrm{N}_{nT}(X'\beta,\ \mathbf{W}\mathbf{\Sigma}_v\mathbf{W}' + I_T \otimes \mathbf{\Sigma}_e)$.

2.2 Model Estimation

Consider a random sample $D_1, \dots, D_K$ of K independent realizations of the total demand vector given by expression (3). In practice, demand may be censored by inventory limits or product availability restrictions imposed by the seller. Unrealized demand is almost always not recorded, and only actual sales data are available for estimation. Denote by $S_1, \dots, S_K$ the corresponding observed sales, with $S_{ki} \leq D_{ki}$ for all $k = 1, \dots, K$ and $i = 1, \dots, nT$. Let $\delta_k \in \{0, 1\}^{nT}$ be the vector of censoring indicators for D_k, defined as $\delta_{ki} = 1$ if $D_{ki} = S_{ki}$ and $\delta_{ki} = 0$ if $D_{ki} > S_{ki}$, for all $i = 1, \dots, nT$. In practice, the censoring indicators are recorded by observing stock or capacity levels. The standard assumption which we make here is that if there is a stock-out or if capacity is fully utilized by sales in one period, then there is potential (unobserved) demand that could not be fulfilled and was censored. In such cases the δ indicator takes the value zero. Otherwise, if there is still stock or capacity available for sales at the end of the period, we assume that the demand has been entirely fulfilled (and observed) and the δ indicator takes the value one.

The problem consists in estimating the parameters of demand model (3) in this classic incomplete data framework, where the latent random shock v is unobservable and the demand realizations $D_1, \dots, D_K$ are potentially censored. The complete but unobserved data are the values of the latent $v_1, \dots, v_K$ and of the uncensored demand $D_1, \dots, D_K$. The observed but incomplete data consists in the sales variables $S_1, \dots, S_K$ and the censoring indicators $\delta_1, \dots, \delta_K$. Based on sales and censoring data $\{S_k, \delta_k\}$, we require estimates of the attribute effects β, of the covariance matrix $\mathbf{\Sigma}_v$ of the latent v, and of the error covariance matrix $\mathbf{\Sigma}_e$.

We estimate the parameters of demand model (3) through maximum likelihood inference. The following proposition gives the expression of the log-likelihood function for the complete data.

Proposition 1 *The logarithm of the likelihood function for the complete data is given by*

$$\log \mathcal{L}(\beta, \mathbf{\Sigma}_v, \mathbf{\Sigma}_e, \{v_k\}) = -(n+1)TK\log(2\pi)/2 - K \cdot (\log \mid \mathbf{\Sigma}_v \mid + T\log \mid \mathbf{\Sigma}_e \mid)/2$$
$$-\frac{1}{2} \cdot \sum_{i=1}^{K} [(D_k - X'\beta - \mathbf{W}v_k)'(I_T \otimes \mathbf{\Sigma}_e^{-1})(D_k - X'\beta - \mathbf{W}v_k) + v_k'\mathbf{\Sigma}_v^{-1}v_k]. \tag{4}$$

Proof. Since $D \sim \mathrm{N}_{nT}(X'\beta,\ \mathbf{W\Sigma}_v\mathbf{W}' + I_T \otimes \mathbf{\Sigma}_e)$, we have that $D_k - X'\beta \sim \mathrm{N}_{nT}(0,\ \mathbf{W\Sigma}_v\mathbf{W}' + I_T \otimes \mathbf{\Sigma}_e)$. Also, $\mathrm{Cov}(v_k, D_k - X'\beta) = \mathbf{\Sigma}_v\mathbf{W}'$. Therefore, the joint distribution of the uncensored demand and of the random shock is multivariate normal,

$$\begin{bmatrix} D_k - X'\beta \\ v_k \end{bmatrix} \sim \mathrm{N}_{(n+1)T}(0, \mathbf{\Sigma}),$$

with covariance matrix given by

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{W\Sigma}_v\mathbf{W}' + I_T \otimes \mathbf{\Sigma}_e & \mathbf{W\Sigma}_v \\ \mathbf{\Sigma}_v\mathbf{W}' & \mathbf{\Sigma}_v \end{bmatrix}. \tag{5}$$

The likelihood function for the complete data is thus

$$\mathcal{L}(\beta, \mathbf{\Sigma}_v, \mathbf{\Sigma}_e, \{v_k\}) = \left(\prod_{k=1}^{K} \frac{1}{\sqrt{(2\pi)^{(n+1)T} \cdot \mid \mathbf{\Sigma} \mid}}\right)$$
$$\times \exp\left\{-\frac{1}{2} \cdot \sum_{i=1}^{K} \begin{bmatrix} D_k - X'\beta \\ v_k \end{bmatrix}' \mathbf{\Sigma}^{-1} \begin{bmatrix} D_k - X'\beta \\ v_k \end{bmatrix}\right\}. \tag{6}$$

From equation (5), using results from Schneider and Barker (1989) it follows that

$$\begin{aligned} \mid \mathbf{\Sigma} \mid &= \mid \mathbf{\Sigma}_v \mid \cdot \mid (\mathbf{W\Sigma}_v\mathbf{W}' + I_T \otimes \mathbf{\Sigma}_e) - (\mathbf{W\Sigma}_v) \cdot \mathbf{\Sigma}_v^{-1} \cdot (\mathbf{\Sigma}_v\mathbf{W}') \mid \\ &= \mid \mathbf{\Sigma}_v \mid \cdot \mid I_T \otimes \mathbf{\Sigma}_e \mid = \mid \mathbf{\Sigma}_v \mid \cdot \mid \mathbf{\Sigma}_e \mid^T, \end{aligned}$$

hence

$$\log \mid \mathbf{\Sigma} \mid = \log \mid \mathbf{\Sigma}_v \mid + T\log \mid \mathbf{\Sigma}_e \mid. \tag{7}$$

At the same time, we have

$$\mathbf{\Sigma}^{-1} = \left[\begin{array}{cc} I_T \otimes \mathbf{\Sigma}_e^{-1} & -(I_T \otimes \mathbf{\Sigma}_e^{-1}) \cdot \mathbf{W} \\ -\mathbf{W}' \cdot (I_T \otimes \mathbf{\Sigma}_e^{-1}) & \mathbf{\Sigma}_v^{-1} + \mathbf{W}' \cdot (I_T \otimes \mathbf{\Sigma}_e^{-1}) \cdot \mathbf{W} \end{array} \right]. \tag{8}$$

Equation (4) now follows after some algebra, taking logarithms of both sides in (6) and replacing $\log \mid \mathbf{\Sigma} \mid$ and $\mathbf{\Sigma}^{-1}$ with expressions (7) and (8). ■

For the rest of this section and in order to simplify notation, we assume that the error terms variances are all equal to σ_e^2, so that $\mathbf{\Sigma}_e = \sigma_e^2 I_n$. This is not a restrictive assumption, as it is straightforward to extend the results to the case of unequal error variances. With this assumption, we have that $I_T \otimes \mathbf{\Sigma}_e^{-1} = I_{nT}/\sigma_e^2$.

Proposition 2 *The values $\widehat{\beta}$, $\widehat{\mathbf{\Sigma}}_v$, $\widehat{\sigma}_e^2$ and $\{\widehat{v}_k\}$ that maximize the complete data log-likelihood function satisfy the following equations:*

$$\widehat{\beta} = \frac{1}{K}(XX')^{-1}X\sum_{k=1}^{K}(D_k - \mathbf{W}\widehat{v}_k) \tag{9}$$

$$\widehat{\mathbf{\Sigma}}_v = \frac{1}{K}\sum_{k=1}^{K}\widehat{v}_k\widehat{v}_k' \tag{10}$$

$$\widehat{\sigma}_e^2 = \frac{1}{KnT}\sum_{k=1}^{K}(D_k - X'\widehat{\beta} - \mathbf{W}\widehat{v}_k)'(D_k - X'\widehat{\beta} - \mathbf{W}\widehat{v}_k) \tag{11}$$

$$\widehat{v}_k = (\mathbf{W}'\mathbf{W} + \widehat{\sigma}_e^2\widehat{\mathbf{\Sigma}}_v^{-1})^{-1}\mathbf{W}'(D_k - X'\widehat{\beta}), \text{ for } k = 1, \ldots, K. \tag{12}$$

Proof. We first derive the score equations by taking partial derivatives of equation (4). We have

$$\frac{\partial \log \mathcal{L}}{\partial \beta} = -\frac{1}{\sigma_e^2}\sum_{k=1}^{K}[\beta' X - (D_k - \mathbf{W}v_k)']X', \tag{13}$$

and equation (9) follows by setting (13) equal to zero. Also,

$$\frac{\partial \log \mathcal{L}}{\partial \mathbf{\Sigma}_v^{-1}} = \frac{K}{2} \cdot \mathbf{\Sigma}_v - \frac{1}{2}\sum_{k=1}^{K} v_k' v_k,$$

hence equation (10) holds. We have

$$\frac{\partial \log \mathcal{L}}{\partial \sigma_e^2} = -\frac{nKT}{2} \cdot \frac{1}{\sigma_e^2} + \frac{1}{2(\sigma_e^2)^2} \sum_{k=1}^{K} (D_k - X'\widehat{\beta} - \mathbf{W}\widehat{v}_k)'(D_k - X'\widehat{\beta} - \mathbf{W}\widehat{v}_k),$$

thus we derive equation (11). Finally,

$$\frac{\partial \log \mathcal{L}}{\partial v_k} = -\frac{1}{\sigma_e^2} \cdot [v_k'\mathbf{W}' - (D_k - X'\beta)']\mathbf{W} - v_k'\mathbf{\Sigma}_v^{-1}, \tag{14}$$

and equation (12) follows by setting (14) equal to zero. ■

The following corollary of Proposition 2 results from first principles.

Corollary 3 *A set of sufficient statistics for* $\boldsymbol{\beta}$, σ_e^2 *and* $\{v_k\}$ *are given by* $\sum_{k=1}^{K}(D_k - \mathbf{W}\widehat{v}_k)$, $\sum_{k=1}^{K}(D_k - X'\widehat{\beta} - \mathbf{W}\widehat{v}_k)'(D_k - X'\widehat{\beta} - \mathbf{W}\widehat{v}_k)$, *and* $\{(\mathbf{W}'\mathbf{W} + \widehat{\sigma}_e^2\widehat{\mathbf{\Sigma}}_v^{-1})^{-1}\mathbf{W}'(D_k - X'\widehat{\beta})\}_k$, *respectively.*

Note that if the complete data would be available so that the uncensored demand $\{D_k\}$ is observable, the log-likelihood function given by (4) could be maximized to compute the maximum likelihood estimates of β, $\mathbf{\Sigma}_v$, and $\mathbf{\Sigma}_e$. We do not, however, observe the complete data, hence direct maximization of the log-likelihood function is not possible. Instead, we estimate the model parameters using the EM algorithm (Dempster, Laird and Rubin 1977), which is the classic approach for maximum likelihood inference with incomplete or missing data. The EM algorithm starts with arbitrary initial estimates and iterates between two steps updating the value of the parameter vector at each iteration. In the expectation (E) step, the algorithm computes expected values of the sufficient statistics for the complete data, conditional on the observed data and on the current values of the parameters. In the maximization (M) step, the likelihood is computed by substituting the missing data with their expected values from the previous E–step, and new estimates of the parameters are obtained by maximizing the likelihood. The algorithm iterates between the E and the M steps until convergence. In practice, different convergence criteria can be used, including a maximum change from one iteration to the next in the value of the log-likelihood function or in the values of the estimated parameters.

Let $\theta = \{\beta, \mathbf{\Sigma}_v, \sigma_e^2, \{v_k\}\}$ be the parameter set, and let $\widehat{\theta}^{(r)}$ be the estimated value of θ after the r–th iteration. The EM algorithm starts with initial estimates $\widehat{\theta}^{(0)}$ which may be obtained using mixed model techniques and treating censored observations as ignorably missing. At the r-th iteration, the E and M steps of the algorithm are given by:

- E step: Compute $E[\log \mathcal{L}(\theta) \mid \widehat{\theta}^{(r-1)}]$.

- M step: Maximize the expected complete data log-likelihood to find

$$\widehat{\theta}^{(r)} = \arg\max_{\theta} E[\log \mathcal{L}(\theta) \mid \widehat{\theta}^{(r-1)}].$$

In the rest of this subsection we describe the details of the E step. At the r-th iteration, the E step computes the expected values of the sufficient statistics for the complete data, conditional on the observed data given by sales and censoring indicators $\{\mathbf{S}_k, \delta_k\}$ and on the parameters $\widehat{\theta}^{(r-1)}$ estimated in the previous iteration.

Let us denote $\widetilde{D}_k^{(r)} = E[D_k \mid S_k, \delta_k, \widehat{\theta}^{(r)}]$, the conditional expected value of the uncensored demand computed at the r-th iteration. Using Corollary 3, the expected values of the sufficient statistics for β, σ_e^2, and $\{v_k\}$ at the r-th iteration are given respectively by

$$\mathrm{E}[\sum_{k=1}^{K}(D_k - \mathbf{W}v_k) \ \mid\ \{S_k, \delta_k\}, \widehat{\theta}^{(r-1)}] = \sum_{k=1}^{K}(\widetilde{D}_k^{(r-1)} - \mathbf{W}v_k^{(r-1)}), \tag{15}$$

$$\begin{aligned} &\mathrm{E}[\sum_{k=1}^{K}(D_k - X'\beta - \mathbf{W}v_k)'(D_k - X'\beta - \mathbf{W}v_k) \ \mid\ \{S_k, \delta_k\}, \widehat{\theta}^{(r-1)}] \qquad (16) \\ &= \sum_{k=1}^{K}\{\mathrm{E}[D_k'D_k \ \mid\ S_k, \delta_k, \widehat{\theta}^{(r-1)}] - 2(X'\beta^{(r-1)} + \mathbf{W}v_k^{(r-1)})'\widetilde{D}_k^{(r-1)} \\ &\quad +(X'\beta^{(r-1)} + \mathbf{W}v_k^{(r-1)})'(X'\beta^{(r-1)} + \mathbf{W}v_k^{(r-1)})\}, \end{aligned}$$

and

$$\begin{aligned} &\mathrm{E}[(\mathbf{W}'\mathbf{W} + \widehat{\sigma}_e^2\widehat{\mathbf{\Sigma}}_v^{-1})^{-1}\mathbf{W}'(D_k - X'\beta) \ \mid\ S_k, \delta_k, \widehat{\theta}^{(r-1)}] \\ &= (\mathbf{W}'\mathbf{W} + \widehat{\sigma}_e^{2(r-1)}\widehat{\mathbf{\Sigma}}_v^{-1(r-1)})^{-1}\mathbf{W}'(\widetilde{D}_k^{(r-1)} - X'\widehat{\beta}^{(r-1)}), \quad k = 1, \ldots, K. \end{aligned} \tag{17}$$

In order to compute these expectations we need to evaluate $\widetilde{D}_k^{(r-1)} = E[D_k \mid S_k, \delta_k, \widehat{\theta}^{(r-1)}]$ and $\mathrm{E}[D_k'D_k \mid S_k, \delta_k, \widehat{\theta}^{(r-1)}]$. Note that we have

$$D_k \mid \widehat{\theta}^{(r-1)} \sim \mathrm{N}_{nT}(X'\beta^{(r-1)} + \mathbf{W}v_k^{(r-1)},\ \sigma_e^{2(r-1)}\mathbf{I}_{nT}),$$

thus the components of D_k are independent and

$$D_{kj} \mid \widehat{\theta}^{(r-1)} \sim N((X'\beta^{(r-1)})_j + (\mathbf{W}v_k^{(r-1)})_j,\ \sigma_e^{2(r-1)}),$$

for $j = 1, \dots, nT$, where we denote by $(A)_j$ the j–th component of vector A.

If $\delta_{kj} = 1$ then D_{kj} is uncensored, hence $D_{kj} = S_{kj}$ and therefore $\widetilde{D}_{kj}^{(r)} = E[D_{kj} \mid S_{kj}, \widehat{\theta}^{(r)}] = S_{kj}$ and $\mathrm{E}[D_{kj}^2 \mid S_{kj}, \widehat{\theta}^{(r)}] = S_{kj}^2$. If $\delta_{kj} = 0$, then D_{kj} is censored and $D_{kj} > S_{kj}$. From standard results for the truncated normal distribution (Johnson et al. (1994), p.156–162), it follows that

$$\widetilde{D}_{kj}^{(r)} = E[D_{kj} \mid D_{kj} > S_{kj}, \widehat{\theta}^{(r)}] = (X'\beta^{(r)})_j + (\mathbf{W}v_k^{(r)})_j + \sigma_e^{(r)} \cdot \frac{\phi(z_{kj}^{(r)})}{1 - \Phi(z_{kj}^{(r)})}, \tag{18}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cumulative density functions, and

$$z_{kj}^{(r)} = \frac{S_{kj} - [(X'\beta^{(r)})_j + (\mathbf{W}v_j^{(r)})_j]}{\sigma_e^{(r)}}.$$

Note also that

$$\mathrm{E}[D_k' D_k \mid S_k, \delta_k, \widehat{\theta}^{(r-1)}] = \sum_{j=1}^{nT} \mathrm{E}[D_{kj}^2 \mid S_k, \delta_k, \widehat{\theta}^{(r-1)}] \tag{19}$$

and

$$\begin{aligned} \mathrm{E}[D_{kj}^2 \mid D_{kj} > S_{kj}, \widehat{\theta}^{(r)}] = {} & [(X'\beta^{(r)})_j + (\mathbf{W}v_k^{(r)})_j]^2 + \sigma_e^{2(r)} \\ & + \ \sigma_e^{(r)} \cdot \frac{\phi(z_{kj}^{(r)})}{1 - \Phi(z_{kj}^{(r)})} \cdot [S_{kj} + (X'\beta^{(r)})_j + (\mathbf{W}v_k^{(r)})_j]. \end{aligned}$$

Replacing the expressions for $\widetilde{D}_{kj}^{(r)}$ and $\mathrm{E}[D_{kj}^2 \mid S_{kj}, \widehat{\theta}^{(r)}]$ in (15)–(17), we obtain the expected values of the sufficient statistics for β, σ_e^2 and $\{v_k\}$. These are then substituted in the log–likelihood function in the M step of the EM algorithm in order to update the estimates of the parameter set $\widehat{\theta}^{(r)}$.

The EM algorithm has the advantage that it converges reliably to the maximum likelihood estimates under certain mild conditions (Wu 1983). In practice, ad-hoc stopping rules are commonly used, such as a maximum change in the log-likelihood or in the estimated parameters from one iteration to the next. A disadvantage of the EM algorithm is that its rate of convergence can be slow, particularly when there are a large number of parameters or when a high percentage of the sample data is censored. In the next section, we investigate the convergence speed of the algorithm under different censoring conditions using a simulation study.

After convergence of the EM algorithm to the maximum likelihood estimates, the standard errors of the estimates may be computed using a bootstrap approach (Efron and Tibshirani, 1998, Chapter 6). We illustrate the use of the bootstrap in the applications described in Section 4.

3 Simulation Experiments

In this section we investigate the performance of the EM algorithm described in Section 2.2 through a series of simulation experiments. The objective of the simulation study is to examine the effects of demand correlation, degree of censoring and length of selling horizon on the properties of parameter estimates and on the speed of convergence of the EM algorithm.

We consider the case when the seller offers $n = 2$ products, with inventory of 40 units for the first product and 160 units for the second. For example, for airline bookings the products may correspond to two fare classes (business and economy) on the same itinerary. The total flight capacity is fixed at 200 seats, with 40 seats allocated to the first fare class and 160 seats allocated to the second.

We simulated demand for the two products over a time horizon with $T = 6,\ 12,$ and 20 selling periods. The mean demand $\beta_t = (\beta_{t,1}, \beta_{t,2})$ is uniform across all periods for each product. The latent shock variances σ^2_{v1} and σ^2_{v2} are chosen such that the coefficient of variation of demand is 0.4, and the correlation between demand for the two products takes values of $\rho = 0, 0.3,$ and 0.6. These choices are typical values in airline demand modelling (McGill 1995; McGill and van Ryzin 1999). We also choose W_t to be the 2×2 identity matrix for each $t = 1, \ldots, T$, and the error variances to be both equal to $\sigma^2_e = 1$.

The simulation comprised 1000 iterations. At each iteration, we generated $K = 500$ realizations of demand from model (3) with parameters specified above. The demand vectors were then censored in order to obtain the sales data, and the percentage of censoring took the values of 0%, 20% and 40%. The censoring indicators for each demand realization were also recorded at this stage. We then used the EM algorithm to estimate the parameters of the demand distribution from the sales data and from the censoring indicators. We assumed that the algorithm has converged when the maximum relative change in one iteration for all estimates was less than 0.001. Tables 1 and 2 summarize the results of the simulations and report respectively the empirical bias and the mean squared error of the parameter estimates.

The bias for the estimated parameters is mostly negative, except for the estimate of the correlation ρ. This shows that the EM algorithm slightly underestimates the true parameter values. The bias and the mean squared error for all parameters are generally decreasing with the percentage of censoring. This simply reflects the fact that more information in the sample naturally leads to better estimates. As expected, the bias and mean squared error for the estimates of the mean demand in each period β_1 and β_2 are not affected by either the length T of the selling horizon, or by the correlation ρ between the components of the latent random effect v. For estimates of the error variance σ_e^2 and of the random effect variances σ_{v1}^2 and σ_{v2}^2, the bias and the mean squared error decrease with the length of the selling horizon. For the estimate of the random effect correlation ρ, the bias and the mean squared error also decrease with T, but they increase with the true value of ρ. The positive impact of the length T of the selling horizon on the properties of the estimates is to be expected, since a longer horizon gives more information on the distribution of the latent v, and thus the estimates should be better.

The results of the simulation study suggest that the estimation approach outlined here works well in large samples and with relatively long time horizons. This is often the case in practice; for example, airlines commonly use booking horizons with up to $T = 24$ snapshots, and have available sales data for hundreds of previous flights for the purpose of demand forecasting.

Problems may sometimes arise in the implementation of the EM algorithm. Our simulation experience has shown that in some instances there are singularities in $\mathbf{\Sigma}_v$ at a given iteration. In this situation, one could use the estimated value of $\mathbf{\Sigma}_v$ from the previous iteration. For the simulation study reported in this section, we encountered non-convergence problems in less than 2% of the iterations. Although the EM algorithm has notoriously slow convergence, it did however converge quite fast in all the simulation scenarios that we considered, for those iterations where convergence has been achieved. The actual convergence speed decreased both with increasing percentage of censorship and with increasing demand correlation.

Table 1: Empirical bias for maximum likelihood estimates computed with the EM algorithm, based on 1000 iterations. For each iteration, $K = 500$ samples of demand were generated under model (3).

	Correlation	Parameter bias					
	ρ	β_1	β_2	σ_{v1}	σ_{v2}	ρ	σ_e
Censoring 0%							
$T = 6$	0	0.001	-0.001	-0.111	-0.110	-0.003	-0.079
	0.3	-0.003	0.000	-0.119	-0.118	0.093	-0.073
	0.6	0.000	-0.002	-0.094	-0.094	0.094	-0.083
$T = 12$	0	-0.001	0.003	-0.049	-0.050	-0.001	-0.043
	0.3	0.003	-0.001	-0.046	-0.047	0.025	-0.043
	0.6	0.002	0.000	-0.050	-0.051	0.063	-0.039
$T = 20$	0	-0.002	-0.003	-0.026	-0.026	0.000	-0.026
	0.3	0.000	0.002	-0.029	-0.029	0.014	-0.026
	0.6	0.001	0.001	-0.030	-0.029	0.028	-0.026
Censoring 20%							
$T = 6$	0	-0.045	-0.043	-0.232	-0.229	0.004	-0.076
	0.3	-0.063	-0.064	-0.230	-0.233	0.255	-0.061
	0.6	-0.032	-0.034	-0.169	-0.166	0.400	0.046
$T = 12$	0	-0.047	-0.047	-0.091	-0.089	0.004	-0.052
	0.3	-0.053	-0.050	-0.089	-0.090	0.053	-0.052
	0.6	-0.054	-0.053	-0.092	-0.093	0.123	-0.045
$T = 20$	0	-0.052	-0.051	-0.049	-0.050	0.002	-0.034
	0.3	-0.055	-0.057	-0.047	-0.049	0.025	-0.033
	0.6	-0.059	-0.058	-0.049	-0.050	0.052	-0.032
Censoring 40%							
$T = 6$	0	-0.173	-0.174	-0.308	-0.302	0.024	-0.112
	0.3	-0.184	-0.184	-0.294	-0.295	0.254	-0.105
	0.6	-0.141	-0.140	-0.234	-0.231	0.386	0.010
$T = 12$	0	-0.073	-0.076	-0.163	-0.164	0.000	-0.074
	0.3	-0.082	-0.080	-0.162	-0.163	0.089	-0.072
	0.6	-0.194	-0.196	-0.122	-0.123	0.141	-0.061
$T = 20$	0	-0.082	-0.081	-0.088	-0.091	0.003	-0.048
	0.3	-0.158	-0.158	-0.079	-0.079	0.048	-0.047
	0.6	-0.178	-0.177	-0.076	-0.077	0.081	-0.045

Table 2: Empirical mean squared error (MSE) for maximum likelihood estimates computed with the EM algorithm, based on 1000 iterations. For each iteration, $K = 500$ samples of demand were generated under model (3).

	Correlation ρ	Parameter MSE					
		β_1	β_2	σ_{v1}	σ_{v2}	ρ	σ_e
Censoring 0%							
$T = 6$	0	0.004	0.004	0.015	0.014	0.004	0.007
	0.3	0.004	0.004	0.017	0.017	0.016	0.006
	0.6	0.004	0.004	0.010	0.011	0.010	0.007
$T = 12$	0	0.004	0.004	0.004	0.004	0.003	0.002
	0.3	0.004	0.004	0.003	0.003	0.003	0.002
	0.6	0.004	0.004	0.004	0.004	0.006	0.002
$T = 20$	0	0.004	0.004	0.002	0.002	0.003	0.001
	0.3	0.004	0.004	0.002	0.002	0.002	0.001
	0.6	0.004	0.004	0.002	0.002	0.002	0.001
Censoring 20%							
$T = 6$	0	0.006	0.006	0.056	0.055	0.009	0.006
	0.3	0.008	0.008	0.055	0.057	0.080	0.004
	0.6	0.005	0.005	0.031	0.030	0.160	0.002
$T = 12$	0	0.006	0.006	0.010	0.009	0.003	0.003
	0.3	0.007	0.006	0.009	0.009	0.006	0.003
	0.6	0.007	0.007	0.010	0.010	0.017	0.002
$T = 20$	0	0.006	0.006	0.003	0.003	0.003	0.001
	0.3	0.007	0.007	0.004	0.004	0.003	0.001
	0.6	0.007	0.007	0.004	0.004	0.004	0.001
Censoring 40%							
$T = 6$	0	0.034	0.034	0.097	0.093	0.010	0.012
	0.3	0.038	0.037	0.089	0.089	0.073	0.011
	0.6	0.024	0.024	0.057	0.057	0.150	0.002
$T = 12$	0	0.010	0.010	0.028	0.028	0.004	0.006
	0.3	0.011	0.010	0.028	0.028	0.011	0.005
	0.6	0.041	0.042	0.016	0.016	0.021	0.004
$T = 20$	0	0.012	0.012	0.009	0.009	0.003	0.002
	0.3	0.031	0.030	0.007	0.007	0.005	0.002
	0.6	0.037	0.037	0.007	0.007	0.008	0.002

4 Applications

In this section we illustrate the modeling methodology developed in Section 2 with the analysis of two booking data sets from the entertainment and the airline industries.

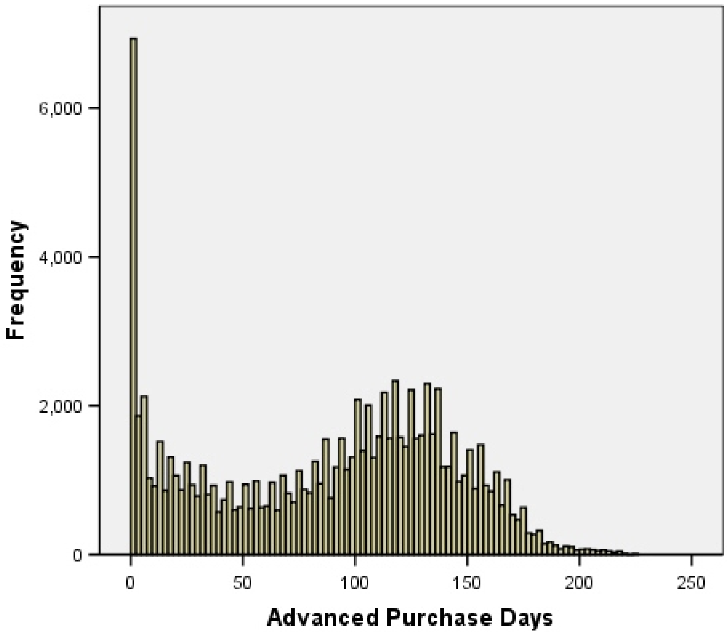
4.1 Theater Booking Data

Our first illustration uses booking data from a major entertainment venue in London. The data set records tickets sold during the 2004–2005 season. There are 21 productions staged on 139 performance dates (hence $K = 139$), with the number of performances for each production varying between three and twelve. We restrict our analysis to evening performances only, and focus on ticket sales for four parts of the house — the Amphitheater, the Balcony, the Grand Tier, and the Orchestra Stalls. The tickets are sold in several different price bands reflecting a decreasing order of prices. In our sample we include sales for four price bands: PB1 contains tickets for Orchestra Stalls and the Grand Tier, PB2 contains tickets for Orchestra Stalls only, PB3 contains tickets for the Balcony, and PB4 contains tickets for the Balcony and the Amphitheater. We define a product as a ticket in one specific price band, therefore we have $n = 4$ products. The corresponding capacity limits in each price band are 490, 177, 112, and 100 seats, and about 15% of the demand observations were censored.

The data set contains a total of 109719 bookings, many of which have been made well in advance of the performance date. Indeed, the number of advanced purchase days varies between 0 and 224, with a mean of 88 days, a median of 98 days, and a standard deviation of 54 days. Figure 4.1 gives the histogram of the highly skewed distribution of days of advanced purchase. An interesting feature of the booking process for theater tickets is that it tends to be bimodal — a significant proportion of bookings are made three to four months before the performance, then bookings decline, and finally there is another surge of last-minute bookings in the week before the performance. For the purpose of our analysis, we aggregate bookings in seven periods during the horizon, hence $T = 7$. Period $t = 1$ is the last week before the performance, the next period $t = 2$ is between one month and one week, and the remaining five periods contain each one month of bookings. This ensures that each period contains between 10%-20% of the total bookings.

Since tickets in all price bands are normally sold during the entire horizon, we take the weighting matrices as the 4×4 identity matrix, $W_t = I_4$ for $t = 1, \ldots, 7$. No product attributes were available, so the covariate vector X_t contains only an intercept and the

Figure 1: Histogram of the distribution of advanced purchase days for theater tickets.



corresponding effect parameter is β_t. We assume that the error variances are equal, hence $\mathbf{\Sigma}_e = \sigma_e^2 I_4$. The parameters to estimate are therefore the mean effects β_t, $t = 1, \ldots, 7$ for each price band, the variance σ_e^2 and the covariance matrix $\mathbf{\Sigma}_v$ of the random shock distribution with components σ_{vi}^2, and σ_{vij}, $i, j = 1, \ldots, 4$.

Table 3 reports the maximum likelihood estimates from fitting demand model (3) to the theater booking data. Since the dynamics of demand for weekend and weekdays performances are quite different, we stratify the data and report the results separately for the subsample of 49 weekend performances and for the subsample of 90 weekdays performances. The EM algorithm converged in less than a minute on each subsample. We assumed that convergence was achieved when the change in all parameter estimates from one iteration to the next was less than 0.001. After convergence, we computed the standard errors of the estimates using the nonparametric bootstrap approach (Efron and Tibshirani, 1998), implemented as follows. We drew a simple random sample with replacement of the available performance days, and for each of the days selected we included in the sample data the corresponding recorded sales and censoring indicators for all four products and all seven booking periods. We then estimated the demand model from the bootstrapped data using the EM algorithm, and repeated the process 500 times resulting in 500 sets of estimates. The bootstrap standard error of each parameter is then computed as the sample standard deviation across these 500 sets of estimates. These standard errors are reported in parentheses in Table 3.

The estimated means β_t reflect the average bookings for each product in each period. For tickets in price band PB1 (the most expensive), β_t generally increases with t as most of the bookings tend to be made well in advance of the performance date. This is true for both weekend and weekdays performances, although for weekdays performances the sequence of β_t has a U-shape pattern; a lot of bookings are made more than four months in advance, then they decrease, and finally they slightly increase again in the weeks before the performance. This is consistent with expectations of customer behaviour; expensive tickets for weekend performances are mostly purchased four to six months in advance. For weekdays performances where it may be more difficult to commit so long in advance, most bookings are clustered either four to six months earlier or in the last few weeks before the performance.

The opposite pattern holds for the less expensive products in price bands PB3 and PB4. Here relatively few bookings are made in earlier periods $t = 6, 7$, and a larger proportion of tickets are purchased in the month before the performance. Note that the values of β_1 for tickets in price bands PB3 and PB4 are very similar between weekend and weekdays performances (11.22 versus 10.98, and 30.23 versus 30.66, respectively), showing that last-minute demand for cheaper tickets does not depend on whether the performance is on a weekend or not.

The estimated variances of the latent shock are all statistically significant, showing that there is significant intertemporal demand dependence for both weekend and weekdays performances. The estimated covariance values are also statistically significant, and lead to large correlation coefficients between the components of the latent shock. This implies that there is substantial inter-product demand dependence. The dependence is higher (up to a correlation coefficient of 0.68) among demand for tickets in price bands PB1 and PB2, likely due to substitutable demand for the most expensive products.

4.2 Airline Booking Data

The second example of an application of our modeling methodology uses booking data from a major airline which operates a hub-and-spoke network. The data records tickets sold for 90 departure days of the same flight on a transatlantic route starting from the airline's hub. Only flights departing between Monday and Thursday were included in the sample, since demand for weekend flights may have different characteristics. A product in this setting is a fare class, and we focus our analysis on bookings for two fare classes ($n = 2$), which we

Table 3: Maximum likelihood estimates for demand model (3) from the theater booking data over $T = 7$ periods, stratified by weekdays and weekends. Bootstrap standard errors based on 500 replications are given in parentheses.

Weekend data (49 performances)							
Mean demand parameters	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Price band PB1	28.88	32.00	37.33	52.65	82.39	94.96	84.37
	(2.41)	(3.24)	(4.71)	(6.52)	(7.12)	(7.11)	(12.28)
Price band PB2	10.67	16.63	23.43	23.14	26.65	25.51	9.75
	(2.07)	(2.72)	(3.62)	(3.27)	(3.26)	(3.63)	(1.92)
Price band PB3	11.22	8.69	8.48	14.75	25.50	21.71	12.95
	(1.38)	(1.20)	(1.05)	(2.11)	(2.39)	(2.69)	(2.85)
Price band PB4	30.23	6.50	9.15	11.54	16.95	16.21	5.15
	(2.22)	(0.97)	(1.38)	(1.56)	(1.55)	(1.97)	(1.23)
Latent shock variances	σ^2_{v1}	σ^2_{v2}	σ^2_{v3}	σ^2_{v4}	Error variance		σ^2_e
	35.65	27.14	3.34	4.94			803.76
	(9.53)	(8.68)	(1.06)	(1.15)			(69.98)
Latent shock covariances	σ_{v12}	σ_{v13}	σ_{v14}	σ_{v23}	σ_{v24}	σ_{v34}	
	20.88	3.66	3.61	3.99	3.28	1.08	
	(3.66)	(1.16)	(1.76)	(1.21)	(1.63)	(0.50)	

Weekdays data (90 performances)							
Mean demand parameters	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Price band PB1	42.23	54.07	35.40	35.13	65.60	71.70	85.56
	(3.22)	(3.87)	(2.53)	(2.55)	(5.14)	(4.17)	(7.49)
Price band PB2	18.83	27.95	18.27	16.12	23.56	17.47	7.27
	(2.07)	(2.78)	(2.19)	(1.59)	(2.80)	(2.38)	(1.44)
Price band PB3	10.98	14.67	14.87	13.06	16.97	16.43	7.82
	(0.90)	(1.45)	(1.44)	(1.16)	(1.60)	(1.80)	(1.57)
Price band PB4	30.66	11.20	10.29	11.60	15.87	11.83	2.69
	(1.58)	(0.76)	(0.88)	(1.10)	(1.45)	(1.43)	(0.63)
Latent shock variances	σ^2_{v1}	σ^2_{v2}	σ^2_{v3}	σ^2_{v4}	Error variance		σ^2_e
	51.51	22.92	7.37	6.32			637.06
	(8.02)	(3.49)	(1.70)	(0.84)			(43.43)
Latent shock covariances	σ_{v12}	σ_{v13}	σ_{v14}	σ_{v23}	σ_{v24}	σ_{v34}	
	19.75	8.83	4.97	9.98	6.60	3.53	
	(4.05)	(2.18)	(1.70)	(2.21)	(1.43)	(0.93)	

Table 4: Descriptive statistics for the airline booking data. Ticket sales are recorded for two fare classes (A and B), six booking periods, and 90 departure days.

	Fare class A						Fare class B					
Booking period	1	2	3	4	5	6	1	2	3	4	5	6
Average	0.76	0.89	0.89	0.53	0.49	0.64	0.61	0.61	0.47	0.30	0.24	0.48
Std dev	1.24	1.77	1.79	1.14	1.05	1.34	0.97	1.50	1.18	0.68	0.64	0.91
Maximum	6	9	8	5	4	7	3	9	8	3	4	5
Minimum	0	0	0	0	0	0	0	0	0	0	0	0

denote by A and B. The sample contains the bookings made during the last week before the departure date. They are recorded during six booking periods ($T = 6$), each period representing one day. The maximum correlation of sales for A and B in the same booking period is 0.47, and the maximum correlation of sales across booking periods is 0.31.

Table 4 reports descriptive statistics for the booking data. The average and the standard deviations of the recorded bookings for fare class A are larger in all periods than the average and standard deviations of bookings for fare class B. Since this is a route in high demand and the data represents bookings close to the departure date, around 70% of the demand observations are censored. Due to the high percentage of censoring in all booking periods, the recorded sales in any specific period for many of the 90 departure days are zero. Therefore, the average sales computed over all departure days are all relatively small. However, the maximum sales are much higher. Since these are recorded in days when there is still capacity available for booking, the maximum sales show that demand is potentially much higher than the average sales based on truncated data would indicate.

4.2.1 Model fitting

We assume that demand for fare classes A and B follows model (3) and we estimate the parameters from the censored sales data using the EM algorithm. Since both classes are normally sold during the entire horizon, we take the weighting matrices as the 2 × 2 identity matrix, $W_t = I_2$ for $t = 1, \ldots, 6$. No product attributes were available, so the covariate vector X_t contains only an intercept and the corresponding effect parameter is β_t. We assume that the error variances are equal, hence $\mathbf{\Sigma}_e = \sigma_e^2 I_2$. The parameters to estimate are therefore the

Table 5: Maximum likelihood estimates for demand model (3), computed with the EM algorithm from the airline booking data over $T = 6$ periods. Bootstrap standard errors based on 500 replications are given in parentheses.

Mean demand parameters	β_1	β_2	β_3	β_4	β_5	β_6
Fare class A	3.75	3.58	3.40	4.15	3.97	3.77
	(0.41)	(0.39)	(0.35)	(0.51)	(0.49)	(0.38)
Fare class B	1.88	1.54	1.69	1.87	2.10	1.90
	(0.26)	(0.24)	(0.22)	(0.33)	(0.31)	(0.22)
Latent shock parameters	σ^2_{v1}	σ_{v12}	σ^2_{v2}	Error variance		σ^2_e
	5.19	2.30	1.32			1.41
	(1.37)	(0.94)	(0.59)			(0.15)

mean effects β_t, $t = 1, \ldots, 6$ for each fare class, the variance σ^2_e and the covariance matrix $\mathbf{\Sigma}_v$ of the random shock distribution with components σ^2_{v1}, σ_{v12}, and σ^2_{v2}.

The EM algorithm converged in 57 iterations which required less than a minute. We assumed that convergence was achieved when the change in all parameter estimates from one iteration to the next was less than 0.001. After convergence, we computed the standard errors again using the nonparametric bootstrap as described in Section 4.1.

Table 5 reports the maximum likelihood estimates for demand model (3) from the airline booking data, with bootstrap standard errors given in parentheses. All parameter estimates are highly statistically significant. In particular, the estimates of the random shock variances and covariance are significantly greater than zero, implying that there is indeed substantial correlation of demand for the two fare classes in any single period, and also correlation of demand for any single fare class across different booking periods. Moreover, $\sigma^2_{v1} = 5.19 > 1.32 = \sigma^2_{v2}$ reflecting the fact that there is generally more variability in the demand for fare class A than for fare class B. The estimated mean demands β_t for both fare classes and in all periods are consistently larger than the average recorded sales in the corresponding time periods reported in Table 4. This is a consequence of the high censoring incidence in the data, and it shows that the expected untruncated demands are much larger than the estimates based on average censored sales data. Also, the estimated β_t value for fare class A is larger than the estimated β_t for fare class B in all periods t, preserving the inequality in the average recorded sales for the two fare classes in each period.

4.2.2 Demand untruncation

In this subsection we compare the predictions of untruncated demand computed with the multivariate multiperiod demand model (3) for the airline data to the predictions computed with three alternative approaches. The goal of this comparison is to assess the impact that accounting for demand dependence has on the untruncated demand predictions.

We consider the following models: Model 1 is the most general form of expression (3) with $n = 2$ and $T = 6$, accounting for both inter-product and intertemporal dependence. Model 2 is a restricted form of expression (3) with $n = 1$ and $T = 6$, allowing for intertemporal dependence but not inter-product dependence. Model 3 is an univariate normal demand model for each fare class in each booking period, and it does not capture either inter-product or intertemporal dependence. Finally, we consider the predictions from the standard nonparametric uncensoring method developed by Kaplan and Meier (1958).

Figure 2 gives the uncensored demand predicted with each of the four methods, for both fare classes and all six booking periods. The values predicted by Model 1 which accounts for both inter-product and inter-temporal correlation are larger than those predicted by Models 2 and 3 which ignore some or all dependence patterns. Predictions from Model 1 are also generally larger than the uncensored values computed with the nonparametric Kaplan-Meier method, except for booking period $t = 5$.

4.2.3 Protection levels with dependent demand

In this subsection we show how, in a revenue management setting, the multivariate multiperiod demand model (3) can influence the computation of protection levels for different fare products. We then assess the resulting impact on revenue through a small simulation study that mimics the booking patterns found in the airline booking data in Section 4.2.1.

We focus on a simple and effective heuristic for computing protection levels developed by Belobaba (1987) for the single-leg problem. This heuristic based on the expected marginal seat revenue (EMSR-b) is widely used in practice. It considers fare classes 1 through n sorted in decreasing order of their revenues $p_1 > \ldots > p_{n-1} > p_n$, and it computes the nested protection levels $b = (b_1, \ldots, b_{n-1})$, where b_i is the capacity to be reserved for all fare classes $1, \ldots, i$. A request for fare class $i + 1$ is then fulfilled only if the remaining capacity is greater than b_i. Brumelle and McGill (1993) showed that this policy is optimal when demands are independent and when demand for low-fare products arrives earlier then demand for high-fare products.

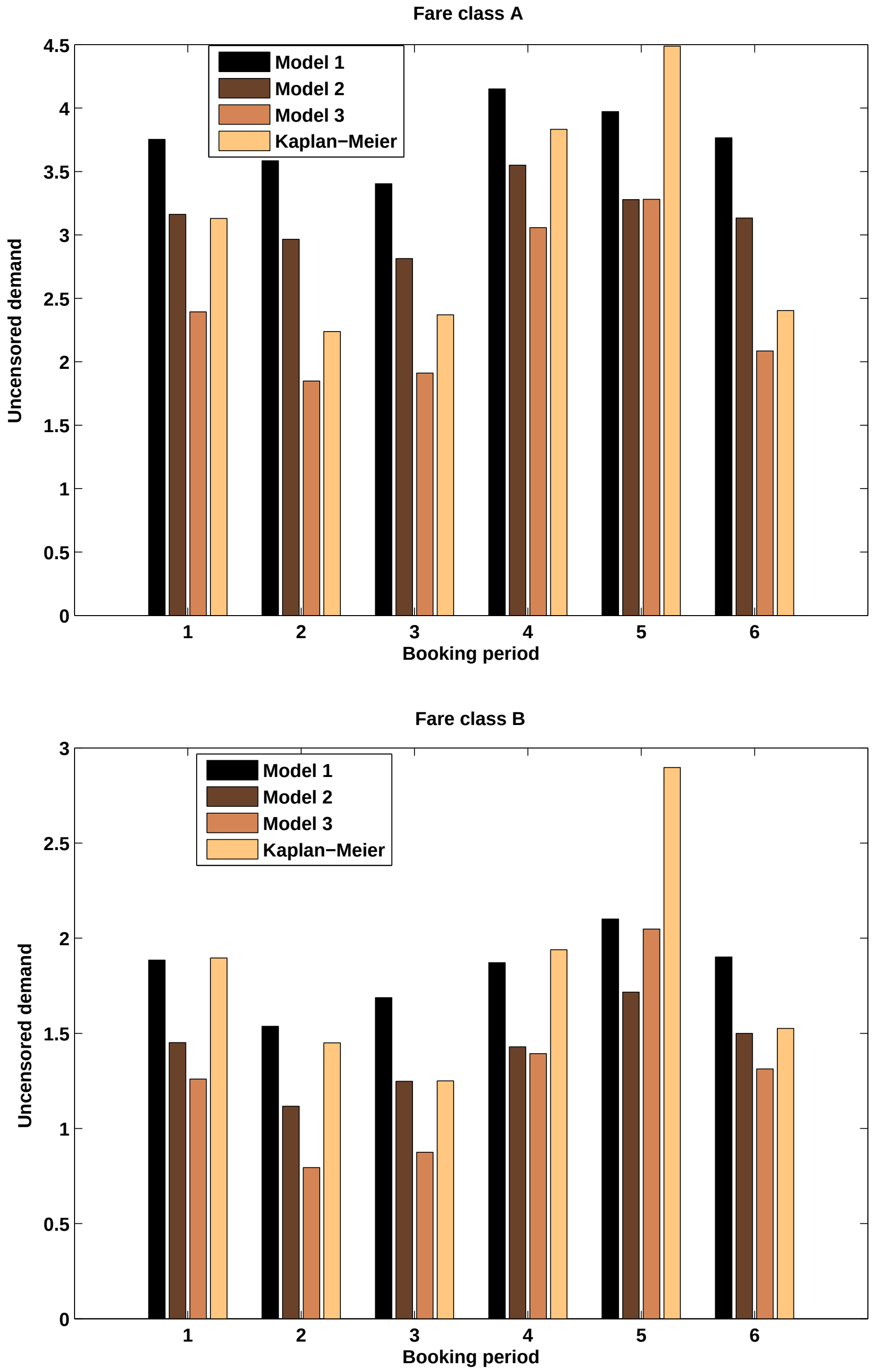


Figure 2: Uncensored demand computed with four different approaches for the airline booking example, for each fare class and each booking period. Model 1 is demand model (3) with $n = 2$ and $T = 6$, allowing for both inter-product and intertemporal dependence. Model 2 is demand model (3) with $n = 1$ and $T = 6$, allowing for intertemporal dependence only. Model 3 is a univariate normal demand model for each fare class in each booking period, without product or time dependence. Kaplan-Meier is the standard nonparametric uncensoring method.

Specifically, let μ_i and σ_i^2 be the mean and variance of demand for fare class i, for all $i \in \{1, \ldots, n\}$. Let $p_i^* = \sum_{j=1}^{i} p_i \mu_i / \sum_{j=1}^{i} \mu_i$ be the weighted average revenue from the first i fare classes. The booking limit b_i is defined as the quantile of the normal distribution that satisfies $p_{i+1} = p_i^* \Pr(X_i^* > b_i)$, where $X_i^* \sim N(\sum_{j=1}^{i} \mu_i, \sum_{j=1}^{i} \sigma_i^2)$ is a normal random variable that intuitively represents the cumulative demand for the first i fare classes.

We focus our comparison on two demand models which we use in order to estimate the means and variances μ_i and σ_i^2 of demand for each fare class. The benchmark is a simple univariate normal demand model for each fare class in each booking period, that does not capture either inter-product or intertemporal demand dependence. The second demand model is the multivariate formulation (3) accounting for both product and serial correlation of demand. Under this model, the estimated means and variances μ_i and σ_i^2 in period t are computed conditional on demand for both products observed up to time t, using the well-known general formulas for conditional means and variances of multivariate normal distributions.[6] Intuitively, the conditional means and variances contain information about past demand and, as demand unfolds over the booking horizon, they convey more information for the computation of protection levels than the estimates based on the univariate demand model that ignores past demand information.

We ran a small simulation study in order to identify the impact on revenue from using the multivariate demand model in the computation of protection levels. We endeavored to mimic the demand patterns from the airline booking data that we analyzed in the previous subsections. For the simulation design we thus chose a setting with two products ($n = 2$) and six booking periods ($T = 6$). We assumed that random product demand exhibits both inter-product and intertemporal dependence, however in order to check the robustness of our methodology we did *not* assume that demand was generated from model (3). Instead, we generated demand from the multivariate normal distribution $\mathrm{N}_{nT}(\beta,\ \mathbf{W}\mathbf{\Sigma}_v\mathbf{W}' + \sigma_e^2 \times I_{nT})$, where $\mathbf{W}$ is the $nT \times n$ matrix obtained by stacking T times the identity matrix I_n, and $\mathbf{\Sigma}_v$, σ_e^2 and β have components given by the estimated values from Table 5. This choice of parameters ensures that random product demand is similar to that recorded in the airline

[6] For any random vectors X_1 and X_2 with a multivariate normal distribution such that

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

we have

$$E[X_1 \mid X_2 = x] = \nu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x - \nu_2),$$

and

$$Var[X_1 \mid X_2 = x] = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Table 6: Percentage revenue gain from using ESMR-b protection levels based on the multivariate rather than on the univariate demand model, under different censoring scenarios.

Censoring percentage	60%	70%	80%	90%
Percentage revenue gain	1.67%	2.99%	3.16%	11.22%

data. We assumed revenues of $p_1 = 1500$ for fare class B and $p_2 = 300$ for fare class A.

We simulated $K = 500$ demand instances for the calibration stage, then a further 100 demand samples for the validation stage. We censored the demand at several censoring levels varying between 60% and 90%, similar to censoring patterns encountered in the airline booking set. We estimated the parameters of the multivariate and univariate demand models using the censored calibration data, then we used the estimated parameters from each model to compute the protection levels. For each of the 100 demand instances in the validation sample, we simulated arrivals uniformly distributed over each booking period, and we allocated seats using the ESMR-b approach with protection levels reoptimized before the allocation decision on each request.

Table 6 summarizes the increase in revenue under different censoring scenarios. The table reports the percentage gain defined as the percentage revenue increase over the 100 demand samples in the validation data, obtained from using protection levels based on the multivariate demand model that accounts for both inter-product and intertemporal correlation rather than on the univariate demand model that ignores both forms of correlation. As expected, the gain in revenue increases with the censoring percentage, as the potential improvement from using a superior method is larger when capacity becomes more scarce. The revenue increases by up to 11% under severe censoring, but even under more moderate censoring levels the gain from using the multivariate demand model can be substantial and of major impact in such a competitive industry.

This example inspired by real booking data shows that protection levels based on statistical demand models that ignore the product and serial dependence of demand fall short of realizing the full revenue potential. Using instead the multivariate demand model for updating protection levels as demand unfolds over time can have a major positive impact on revenue, even in a setting with only two products and a relatively simple heuristic. We expect that this impact is even larger in the case of more than two products or of longer booking horizons.

5 Conclusions

Much research effort has been devoted to building models and developing algorithms to improve operational decisions in diverse fields such as inventory control, supply chain management and revenue management. What often stands between models and practice is the population of model parameters. Accurately estimating parameters and fitting models to data can make a big difference in revenues and an even bigger difference in profits for many firms. Yet the problem of model calibration from data and of parameter estimation is usually not addressed in research papers that focus generally on the stochastic or algorithmic aspects of the models. This paper aims to bridge the gap between the theoretical development of models and their implementation in operations applications that require multivariate demand formulations. We show how improved demand estimates lead to better decisions and how statistical calibration is key in deploying and increasing the impact of models and algorithms.

We have investigated in this paper a class of multivariate demand models that can account for both time dependence and product dependence of demand. These dependence features occur often in practice in a wide range of applications, most commonly arising from substitution and cross–selling effects, or from censoring due to capacity constraints. It is therefore important to capture these demand dependence patterns in order to obtain unbiased and efficient estimates of future demand. The class of models that we investigate here have the flexibility to account for a range of dependence patterns through the inclusion of the common multivariate latent shock. The models could also be extended to include a stochastic autocorrelated process (for example a mean-reverting process) for the latent v_t.

We have also developed in this paper a methodology for estimating the parameters of multivariate and multi-period demand models from censored sales data. We have shown through simulation experiments that our approach based on the EM algorithm is computationally attractive, and that it leads to maximum likelihood estimates with good properties under a range of demand and censoring scenarios. Although the convergence of the EM algorithm is notoriously slow, in our numerical experience with simulations and with practical examples the algorithm has converged relatively fast.

We have illustrated our methodology with the analysis of two booking data sets from the entertainment and the airline industries. We showed that there is strong evidence of significant intertemporal and interproduct correlation in bookings of theater tickets, and uncovered the dynamics of the booking process during the sales horizon for tickets in different price bands. For the airline booking data, we also showed that the uncensored demand predicted by the multivariate models is generally higher than the uncensored values predicted by other models which ignore some or all patterns of dependence. In addition, we exemplified the use of the multivariate demand models for computing protection levels in a revenue management setting, and found that they lead to a substantial increase in revenue relative to the use of protection levels based on univariate demand models.

Among other applications, the demand models and estimation methodology investigated in this paper are useful for the development of inventory management methods for retailing. In a setting with multiple substitutable products, it is important for manufacturers and retailers to assess how the absence of one product affects the demand for other similar products. For example, due to space restrictions, the Macy's department store chain offers collections with fewer color variants in small stores than in larger stores. It is then critical to determine the subset of products that will sell best and the optimal inventory levels, and the models discussed in this paper can help achieve this.

The methodology that we developed here can also be used for revenue management purposes. We outlined in Section 4.2.3 the use of the multivariate demand models for computing protection levels and showed that this leads to a substantial increase in revenue even with a simple protection level policy and in a two-product setting. These demand models have also been used in connection with multistage stochastic programming in order to compute optimal product prices — DeMiguel and Mishra (2006) give an early example of their application in connection with pricing. The practical implementation of the multivariate and multi-period demand models in other application areas and the implications of their use for pricing purposes are potential topics of future research.

Acknowledgment

We are grateful to Victor de Miguel, Guillermo Gallego and Bruce Hardie for their comments and suggestions at various stages of preparation of this manuscript. The support of an RAMD grant from London Business School is gratefully acknowledged.

References

Agrawal, N., S.A. Smith 1996. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics* **43** 839–861.

Anupindi, R., M. Dada, S. Gupta 1998. Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science* **17** 406–423.

Belobaba, P.P. 1987. *Air travel demand and airline seat inventory management.* Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Brumelle, S.L., J.I. McGill 1993. Airline seat allocation with multiple nested fare classes. *Operations Research* **41** 127–137.

Cachon, G.P., C. Terwiesch, Y. Xu 2005. Retail assortment planning in the presence of consumer search. *Manufacturing and Service Operations Management* **7** 330–346.

Conlon, C.T., J.H. Mortimer 2007. Demand estimation under incomplete product availability. Working paper, Harvard University.

Cooper, W.L., Homem-de-Mello, T., A.J. Kleywegt 2006. Models of the spiral-down effect in revenue management. *Operations Research* **54** 968–987.

DeMiguel, V., N. Mishra 2006. What multistage stochastic programming can do for network revenue management. Working paper, London Business School.

Dempster, A.P., N.M. Laird, D.B. Rubin 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B* **39** 1–38.

Efron, B., R.J. Tibshirani 1998. *An Introduction to the Bootstrap*, Chapman and Hall, Boca Raton.

Johnson, N.L., S. Kotz, N. Balakrishnan 1994. *Continuous Univariate Distributions. Volume 1*, 2nd Edition, Wiley, New York.

Kaplan, E.L., P. Meier 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53** 457–481.

Klein, J.P., M.L. Moeschberger 2005. *Survival Analysis*, Springer, New York.

Kök, A.G., M.L. Fisher 2007. Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research* **55** 1001–1021.

Koschat, M.A 2008. Store inventory *can* affect demand: Empirical evidence from magazine retailing. *Journal of Retailing* **84** 165–179.

Lariviere, M., E.L. Porteus 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science* **45** 346–358.

Lau, H.S., A.H. Lau 1996. Estimating the demand distributions of single-period items having frequent stockouts. *European Journal of Operational Research* **92** 254–265.

Lawless, J.F. 2003. *Statistical Models and Methods for Lifetime Data*, Wiley, New York.

McGill, J.I 1995. Censored regression analysis of multiclass passenger demand data subject to joint capacity constraints. *Annals of Operations Research* **60** 209–240.

McGill, J.I., G.J. van Ryzin 1999. Revenue management: Research overview and prospects. *Transportation Science* **33** 233–256.

Nahmias, S. 1994. Demand estimation in lost sales inventory systems. *Naval Research Logistics* **41** 739–757.

Queenan, C.C., M. Ferguson, J. Higbie, R. Kapoor 2007. A comparison of unconstraining methods to improve revenue management systems. *Production and Operations Management* **16** 729–746.

Ratliff, R.M., B.V. Rao, C.P. Narayan, K. Yellepeddi 2008. A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management* **7** 153–171.

Rossi, P.E., G. Allenby, R. McCulloch 2006. *Bayesian Statistics and Marketing*. Wiley, New York.

Schneider, H., G.P. Barker 1989. *Matrices and Linear Algebra*. Dover, New York.

Talluri, K., G.J. van Ryzin 2003. *The Theory and Practice of Revenue Management*. Kluwer Academic Press.

Talluri, K., G.J. van Ryzin 2004. Revenue management under a general discrete choice model of consumer behaviour. *Management Science* **50** 15–33.

Tan, B., S. Karabati 2004. Can the desired service level be achieved when the demand and lost sales are unobserved? *IIE Transactions* **36** 345–358.

van Ryzin, G. 2005. Future of revenue management: Models of demand. *Journal of Revenue and Pricing Management* **4** 204–210.

van Ryzin, G., and S. Mahajan 1999. On the relationship between inventory costs and variety benefits in retail assortments. *Management Science* **45** 1496–1509.

van Ryzin, G., and J. McGill 2000. Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Science* **50** 15–33.

Vulcano, G., G. van Ryzin, R. Ratliff 2008. Estimating primary demand for substitutable products from sales transaction data. Working paper, Columbia Business School.

Weatherford, L.R., and S. Pölt 2002. Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues. *Journal of Revenue and Pricing Management* **1** 234–254.

Wecker, W.E. 1978. Predicting demand from sales data in the presence of stockouts. *Management Science* **24** 1043–1054.

Wu, C.F. 1983. On the convergence properties of the EM algorithm. *The Annals of Statistics* **11** 95–103.