Forecasting Demand for Fashion Goods: A Hierarchical Bayesian Approach

Phillip M. Yelland

Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, U.S.A.
phillip.yelland@gmail.com

Xiaojing Dong

Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053, U.S.A.
xdong1@scu.edu

Abstract

A central feature of demand for products in the fashion apparel segment is a pronounced *product life cycle*—demand for a fashion product tends to rise and fall dramatically in accordance with the rate of public of adoption. Product demands that vary in such a manner can be difficult to forecast, especially in the critical early period of a product's life, when observed demand can be a very unreliable yardstick of demand later on. This paper examines the applicability of a Bayesian forecasting model—based on one developed for use in the computer industry—to fashion products. To do so, we use an agent-based simulation to produce a collection of demand series consistent with commonly-accepted characteristics of fashion adoption. Using Markov chain Monte Carlo techniques to make predictions using the Bayesian model, we are able quantitatively to demonstrate its superior performance in this application.

1. Introduction

According to the US Office of Technology Assessment (1987), the apparel market in the US can be divided into three segments: 1) *fashion* products, which have a very short product life cycle, around ten weeks, and account for approximately 35% of the market; 2) *seasonal* products, which have a slightly longer product life cycle of around twenty weeks and account for approximately 45% of the market; and 3) *basic* products, which do not have an obvious sales pattern and are sold throughout the year. The fashion clothing segment presents a particular challenge for demand planners: Competition is frequently fierce and profit margins volatile, supply chains span the globe and can be hard to coordinate, and the nature of demand for fashion goods themselves is (*ipso facto*) almost invariably highly labile and difficult to predict (Hunter and Valentino 1995). Regarding the latter, the most salient determinant of demand for fashion goods is their very pronounced product life cycle. Indeed, as Sproles (1981) points out, "[f]ashions are, by definition, temporary cyclical phenomena adopted by consumers for a particular time and situation..., having stages of introduction and adoption by fashion leaders, increasing public acceptance (growth), mass conformity (maturation), and the inevitable decline and obsolescence awaiting all fashions".

This inherent variability of fashion product demand makes forecasting very difficult, especially early in a product's life cycle, when little is known about the product's ultimate level of market acceptance; demand early in a product's life may give scant indication of demand later on in the life cycle. To quote Kang (1999): "[T]he best-selling item of the last season could be the worst-selling item for the coming season". Compounding difficulties is the tendency of fashion cycles to shorten in recent years, a phenomenon Kang (*ibid.*) attributes to advances in media technology and the widespread availability of fashion clothing items.

Of course, appreciably short product lifecycles are not confined to the clothing industry; (Yelland 2010) describes a Bayesian forecasting model developed for computer components, where rapid technological development also leads to very pronounced short lifecycles. We set out to examine the suitability of a similar

Bayesian forecasting model for demands in the fashion apparel segment, hypothesizing that a model which had proven effective in accomodating short product lifecycles held promise for fashion clothing forecasting, too.

In (Yelland 2010), assessment of the forecasting model is based on actual demand experienced by a particular manufacturer (Sun Microsystems Inc.); the use of actual series was largely a consequence of the model's use by that manufacturer. While the use of manufacturer data in that paper provides the assurance that the model has real-world applicability, in the strictest sense it establishes efficacy only for one particular firm over a particular period of time. To broaden the scope of the model in this work, we adopt a different approach: We *simulate* plausible demand series for fashion goods options under a variety of circumstances, and test the accuracy of a model derived from (Yelland 2010) in extrapolating them. Of course, very little would be achieved if the simulation were constructed along the same lines as the forecasting model itself, merely embodying the parameters and latent constructs used in the model. As Cook et al. (2006) demonstrate, such an exercise can play a valuable rôle in validating the correctness of a model implementation, but it would hardly establish the broad applicability of the model itself. Instead, we use an *agent-based simulation* to calculate demand based upon individual purchase decisions, with simulation elements designed to drive those individual decisions in a manner consistent with paradigmatic characteristics of fashion goods and their markets.

In the interests of expositional clarity, we make to simplifications in both the simulation and forecasting model used in this chapter, concentrating predominantly on the treatment of the product life cycle. Thus regular cyclical or "seasonal" effects on demand, for example, are represented only in very rudimentary form; simulation of more complex seasonality represents no challenge, and many models of seasonality from the forecasting literature (such as seasonal dummies) may be added to the forecasting model without difficulty—see Harvey (1989) or Brockwell and Davis (2002) for examples.

The plan of the chapter is as follows: In the next section, we outline related work in the literature. The following section describes in detail the simulation used to generate option demands. We then introduce the forecasting model, with brief prefatory remarks about the Bayesian approach to forecasting in general. The forecasting performance of the model in respect of the simulated option demands is then appraised in comparison with that of a sophisticated benchmark forecasting method. We finish with concluding remarks.

2. Related Work

As Hines and Bruce (2007) recount, the modern fashion apparel industry is characterized by increasingly short product life cycles, which contrast with long lead times and complicated supply chain (typically, it takes two years from the design and one year from production before a garment is sold). This has led to a a proliferation of literature studying supply chain management (Donohue 2000), assortment planning (Caro and Gallien 2007), and—most saliently for our purposes—sales forecasting. The latter features studies using various probability distributions to fit the sales data and forecast the demand—Agrawal and Smith (1996), for example, fit a negative binomial distribution to retail sales data. In the model recounted here, we supplement a probability distribution for sales with an explicit account of the product life cycle. This has many precedents in forecasting models for technology products—c.f. (Wu et al. 2010), among many others.

We use a Bayesian approach in formulating our model to tackle another challenge often faced by forecasters in the fashion apparel industry: A very limited amount of available calibration data. Other researchers in the area have also adopted a Bayesian approach, mostly to account for the dynamic nature of demand updates as more data become available. Early inventory models that allow demand updating using a Bayesian approach include those of Dvoretzky et al. (1952), Scarf (1959), Iglehart (1964), Murray and Silver (1966), Azoury and Miller (1984), Azoury (1985), and Miller (1986). More recently, Iyer and Bergen (1997) model a Quick Response apparel supply contract, allowing the demand distribution to be updated in a Bayesian fashion using early sales data. The demand process in this model is assumed to follow a normal distribution. Other recent studies adopt a two-stage demand process for a fashion product. Gurnani and Tang (1999), for example, allow a retailer to place two orders before the selling season. They derive an optimal ordering policy, while allowing the first-stage demand to be updated via a Bayesian method. In a similar setup, Choi et al. (2003) use a newsvendor model in a two-stage context using Bayesian information updating. Other examples include (Eppen and Iyer 1997) and (Choi et al. 2006), to name a few.

Besides information updating, the model in this chapter illustrates that the Bayesian approach can also be applied using a *hierarchical prior*, so as to obtain forecasts for a single item with only a small amount of historical sales information by "borrowing" information from similar items. This feature of the Bayesian approach is widely adopted in the Marketing and Economics literature (see Ansari et al. (2000), Narayanan and Manchanda (2009), and Dong et al. (2009), for example). A seminal paper in Marketing is (Rossi et al. 1996), where a hierarchical Bayesian framework is adopted to analyze model parameters that are related to individual customers' responses to shopping coupons. In this study, each customer makes a limited number of observed purchases, and individual level inference is not feasible using classical statistics. The hierarchical Bayesian approach, on the other hand, calculates individual level inferences by combining population level parameter estimates with the individual level data—essentially borrowing information from the rest of the population. Rossi et al. (2005) present a detailed discussion of hierarchical Bayesian methods.

As indicated in section 1, to apply our method in an empirical context, we adopt an agent-based approach to simulate demand data. In general methodological terms, agent-based simulation investigates aggregate level phenomena by simulating the behaviors of individual agents (Rand and Rust 2011). This approach allows a proper account to be made of complexity in the data generation process while also being tractable. It has been widely adopted across multiple fields, including organizational science (Cohen et al. 1972), supply chain management (Walsh and Wellman 1999) and Marketing (Goldenberg et al. 2009). North et al. (2010), given an account of Procter & Gamble's successful of application agent-based simulated to improve revenue. Related to our study, agent-based simulation has also be applied to investigate the diffusion of innovation—c.f. for example Schwoon (2006), Delre et al. (2007), Schenk et al. (2007), and Kiesling et al. (2009). Our own agent-based simulation is described in the next section.

3. Data Simulation

[1]

Our simulation of demand for fashion products over time is based on the parallels—noted by Sproles (1981) among others—between fashion goods lifecycles and the more general process of innovation adproduct. To this end, we adapt Rand and Rust's (2011) simulation of innovation adproduct. This starts with a set of *agents*, indexed $1, \ldots, N$, whose communication patterns take the form of a so-called *small-world network* after Watts and Strogatz (1998). The network fixes for each agent i a set of *neighbors*, $\texttt{Neighbors}_i$, comprising those agents which are in direct communication with agent i. We simulate the actions of this set of agents over periods $t = 1, \ldots, T$.[2] In each period t, agent i may elect to purchase the product—an event we denote $\texttt{purch}_{it}$. In the interests of simplicity, we assume that each agent purchases the product once. Furthermore, only a subset of the agents—the product's "potential market", $\texttt{Market} \subseteq \{1, \ldots, N\}$—will *ever* purchase the product. In each period, therefore, agent i's probability of purchase is non-zero only if $i \in \texttt{Market}$ and $i \notin \texttt{Purch}_{t-1}$.[3] We abbreviate this conjunction—informally, agent i's membership of the product's potential market in period t—as $\texttt{pot}_{it}$. When the simulation begins, each agent enters into $\texttt{Market}$ independently,[4] each with the same product-specific probability r. Following a paradigm expressed in models of innovation diffusion dating from (at least) Bass (1969), we assume that such purchases are motivated by either: *a*) an *exogenous* influence, which stems from the promotional activities through mass-marketing channels, such as advertising, or *b*) an *endogenous* influence, resulting from communication with neighboring agents who have already purchased the product. The event $\texttt{purch-ex}_{it}$ denotes agent i's *ex*ogenously-motivated purchase in period t, and $\texttt{purch-end}_{it}$ similarly denotes an *end*ogenously-motivated purchase.

[1]A full summary of the notation in the paper, and a description of the probability distributions used is provided in the appendices.

[2]For concreteness's sake, we will frequently regard one period as one day, though we use the term "period" throughout to emphasize the generality of the framework.

[3]For technical convenience, we take $\texttt{Purch}_0$ to be the empty set.

[4]Throughout this section, events are assumed (conditionally) independent unless the contrary is noted.

Since a purchase may arise exogenously or endogenously, the event $\texttt{purch}_{it}$ is the conjunction of these events, so that $\texttt{purch}_{it} = \texttt{purch-ex}_{it} \vee \texttt{purch-end}_{it}$.

Provided that agent i is in the product's potential market in period t, the probability of s/he making an exogenously-motivated purchase in period t is determined by a parameter p_i associated with agent i and a seasonal effect associated with t. As adumbrated in section 1, to keep things simple, the simulation incorporates only one seasonal effect—a "day-of-the-week" effect that distinguishes every 7^{th} period with a 50% increase in sales. Thus we have:

$$p(\texttt{purch-ex}_{it}|\texttt{pot}_{it}) = \min[(1 + 0.5\texttt{S}_t)p_i, 1], \tag{1}$$

where $\texttt{S}_t$ is an indicator variable picking out every 7^{th} period:

$$\texttt{S}_t = \begin{cases} 1 & \text{if } t \bmod 7 = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The probability of an endogenously-motivated purchase by agent i in period t (conditional on membership of the potential market in t) depends on another agent-specific parameter, q_i, along with the seasonal effect and the fraction of agent i's neighbors who have already purchased the product:

$$p(\texttt{purch-end}_{it}|\texttt{pot}_{it}) = \min\left[(1 + 0.5\texttt{S}_t)q_i \frac{|\texttt{Neighbors}_i \cap \texttt{Purch}_{t-1}|}{|\texttt{Neighbors}_i|}, 1\right]. \tag{3}$$

At the outset of the simulation for a particular product, parameters p_i and q_i for each agent i are sampled from the (compound) beta distributions listed below. We use an alternate (and arguably more intuitive) parameterization, $\text{Beta}(\mu, \sigma^2)$, of the beta distribution in terms of its mean μ and variance σ^2, where it is required that $\sigma^2 < \mu(1-\mu)$; the more conventional parameters α and β can be recovered as $\mu\nu$ and $(1-\mu)\nu$ resp., for $\nu = \mu(1-\mu)/\sigma^2 - 1$. In addition, we use the shorthand $x \sim \{v_1, v_2, \ldots, v_n\}$ to indicate that value of the random variable x is drawn randomly from the set $\{v_1, v_2, \ldots, v_n\}$. Then:

$$p_i \sim \text{Beta}(\mu_p, \mu_p/2), \text{ where } \mu_p \sim \{0.00035, 0.0005, 0.001, 0.005, 0.01, 0.02\}, \tag{4}$$
$$q_i \sim \text{Beta}(\mu_q, \mu_q/2), \text{ where } \mu_q \sim \{0.05, 0.1, 0.2, 0.4, 0.5\}, \tag{5}$$

Simulating a demand series over periods $1, \ldots, T$ involves the following:

1. Initialize the agent-specific values p_i and q_i according to the distributions given in (4) and (5) resp. Compute $\texttt{Market}$ according to r, set $\texttt{Purch}_0$ to $\emptyset$ and $\texttt{pot}_{i1}$ iff $i \in \texttt{Market}$.

   For each time period t:

2. Independently for each agent i such that $\texttt{pot}_{it}$:[5]
   (a) Draw a uniformly-distributed random variable $\texttt{u-exo}_{it} \sim \text{Unif}(0, 1)$. Assert event $\texttt{purch-ex}_{it}$ iff $\texttt{u-exo}_{it} < p(\texttt{purch-ex}_{it}|\texttt{pot}_{it})$, the latter term being defined in equation (1).
   (b) Draw another random variable $\texttt{u-endo}_{it} \sim \text{Unif}(0, 1)$ and assert $\texttt{purch-end}_{it}$ iff $\texttt{u-endo}_{it} < p(\texttt{purch-end}_{it}|\texttt{pot}_{it})$ from equation (3).
   (c) Assert $\texttt{purch}_{it}$ iff $\texttt{purch-ex}_{it}$ or $\texttt{purch-end}_{it}$ is true.

3. Let $\texttt{Purch}_{t+1} = \texttt{Purch}_t \cup \{i \in 1, \ldots, N \mid \texttt{purch}_{it}\}$.

4. Simulated demand, y_t, for the product in period t, is $\sum_{i=1}^{N} \mathcal{I}(\texttt{purch}_{it})$, where $\mathcal{I}(\texttt{purch}_{it})$ is 1 if $\texttt{purch}_{it}$ is true, and 0 otherwise.

[5] i.e. $i \in \texttt{Market}$ and $i \notin \texttt{Purch}_{t-1}$.

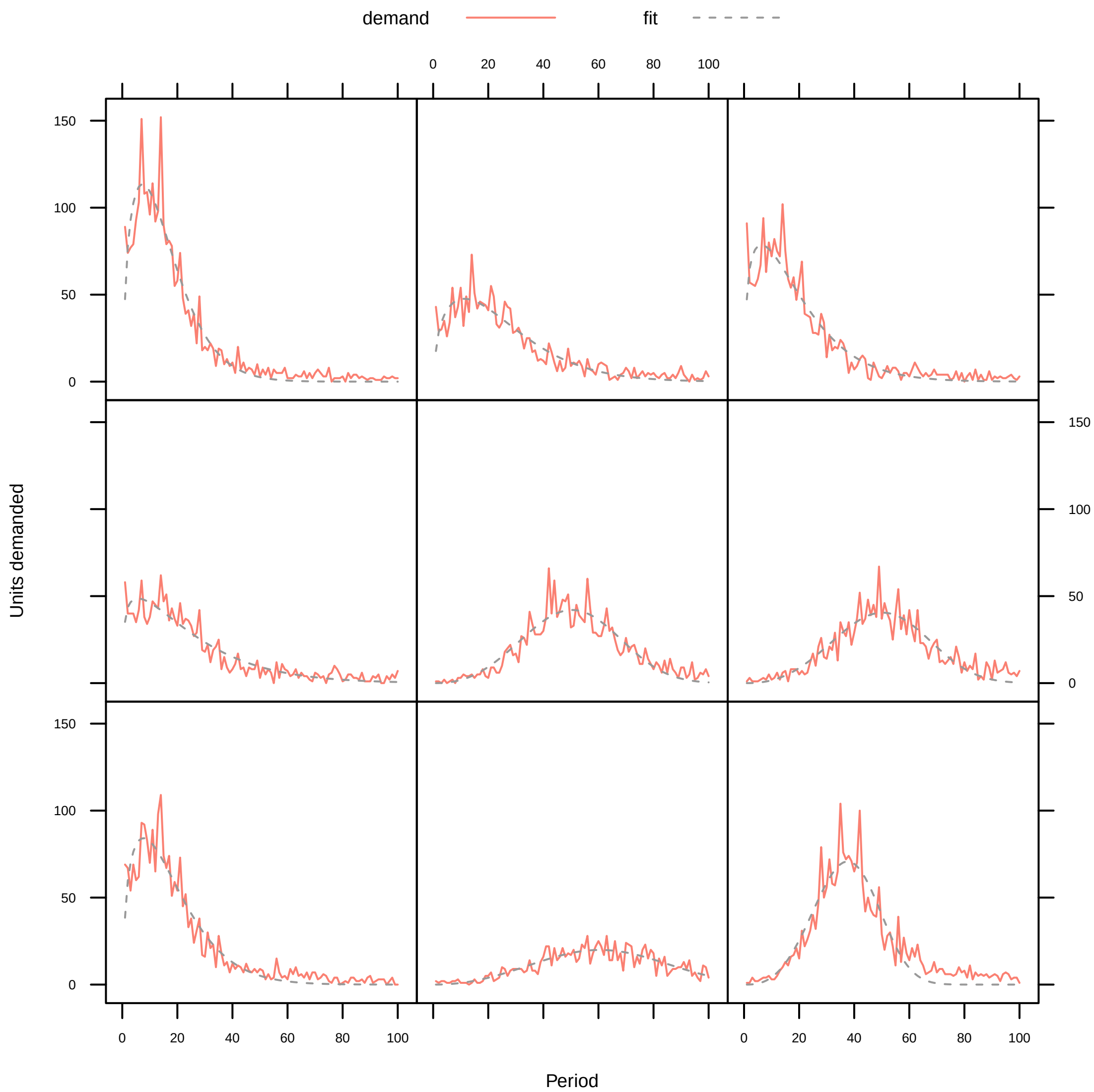

Figure 1: Sample simulated product demands (solid lines), and corresponding fitted Weibull curves (dotted lines). Note the variability of the overall shape of the demand series and of the peak demand in each case.

Repeating the simulation procedure a number of times produces a collection of possible trajectories for product demand over time. We took a sample of 50 such series, each of 100 periods in length, for use in our forecasting exercises; nine typical sample members appear in figure 1. In next section, we describe the Bayesian model proposed for extrapolating these series.

4. Forecasting Model

Before describing the details of forecasting model itself, we outline in general terms how Bayesian methods are particularly appropriate when forecasting demand for fashion goods.

4.1. Bayesian Forecasting

In the conventional forecasting situation examined in the literature, a suitably long time series of observed values—$\boldsymbol{y} = (y_t, \ldots, y_T)$, say—is presented for extrapolation, and (restricting the discussion to a single period forecast horizon for simplicity) the task of the forecaster is to predict the value of the value of the series after h further periods, y_{T+h}. In the demand forecasting exercise examined here, however, short product life cycles mean that individual products frequently lack sufficient observed demand values to support reliable extrapolation. In this application, therefore, input to the forecasting process consists not only of previous demands for a particular product, but also of observed demands for *other* similar products, too—even products that are no longer available. Thus the data takes the form of a collection of series, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J$, of potentially differing lengths (some of which may be zero), so that for $j = 1, \ldots, J$, $\boldsymbol{y}_j = (y_{j1}, \ldots, y_{jT_j})$. The aim is to forecast $y_{l\,T_l+h}$, for some chosen product l.

The objective of the model presented in this section is a statistical representation of such a collection of demand series using a set of unobserved quantities. This latter set—which we denote in the abstract by the vector $\boldsymbol{\theta}$—contains the model parameters, and may also contain the values of latent variables or processes. It is assumed that the representation in terms of $\boldsymbol{\theta}$ is sufficiently detailed that all the elements of the series are conditionally independent given $\boldsymbol{\theta}$, i.e.:[6]

$$p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J|\boldsymbol{\theta}) = \prod_{j=1}^{J} \prod_{t=1}^{T_j} p(y_{jt}|\boldsymbol{\theta}). \tag{6}$$

A Bayesian forecast of $y_{l\,T_l+h}$ rests on its *posterior predictive distribution.* The latter is simply the conditional distribution of $y_{l\,T_l+h}$ given the historical demands, $p(y_{l\,T_l+h}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J)$. The conditional independence property of the model expressed in equation (6) is pivotal to the derivation of this distribution, since on the assumption that $y_{l\,T_l+h}$ is also well-represented by $\boldsymbol{\theta}$, it should be conditionally independent of the historical demands, just as the historical demands were conditionally independent of each other:

$$p(y_{l\,T_l+h}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_J|\boldsymbol{\theta}) = p(y_{l\,T_l+h}|\boldsymbol{\theta})p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J|\boldsymbol{\theta}). \tag{7}$$

Now it is easy to show that with the provision of a prior distribution for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$, the posterior predictive distribution may be expressed as:

$$p(y_{l\,T_l+h}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J) = \int p(y_{l\,T_l+h}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J)d\boldsymbol{\theta} \tag{8}$$

$$\propto \int p(y_{l\,T_l+h}|\boldsymbol{\theta})p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{9}$$

Note that the second factor in the integral on the right hand side of equation (8) is the posterior distribution of $\boldsymbol{\theta}$ given the observed data, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J$. As many treatments of Bayesian statistics illustrate (Bernardo and Smith 1994, Gelman et al. 2003, for example), provided that the observed data is sufficiently informative,

[6] Since some of the T_j may be zero, we adopt the convention that for any expression $\bullet$, $\prod_{j=1}^{0} \bullet = 1$.

even if $p(\boldsymbol{\theta})$ is diffuse or non-informative for elements of $\boldsymbol{\theta}$, the posterior distribution will be sharp enough to yield reasonably precise predictions for $y_{l\,T_l+h}$ in equation (8). This is an advantage in applications such as this, where very little prior information is available in advance of the model's deployment.

We also note that given a mechanism—such as the Markov chain Monte Carlo (*MCMC*) simulator described in Section 5—for drawing samples from the posterior distribution of $\boldsymbol{\theta}$, the right hand side of equation (8) shows how one may sample from the posterior predictive distribution by drawing a value $\tilde{\boldsymbol{\theta}}$ from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J)$ and then drawing one from the conditional distribution $p(y_{l\,T_l+h}|\tilde{\boldsymbol{\theta}})$. The resulting sample may then be used to characterize the posterior predictive distribution for $y_{l\,T_l+h}$ in equation (8), yielding (amongst other quantities) a point forecast for $y_{l\,T_l+h}$.

To allow for some variation in demand patterns between products, we adopt a so-called *hierarchical* or *multilevel* prior (Gelman 2006, Gelman and Hill 2006) in the model. This may be thought of abstractly as dividing $\boldsymbol{\theta}$ into two collections of sub-vectors: A collection $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_J)$ of parameter vectors associated with products, and a single vector $\boldsymbol{\vartheta}$ of common "population-level" parameters. The prior for $\boldsymbol{\theta}$ as a whole is expressed by defining the priors for the product-level parameters in terms of the values of the population-level parameters, while the population-level parameters receive their own free-standing priors. This means that in equation (9):

$$p(\boldsymbol{\theta}) = \left\{ \prod_{j=1}^{J} p(\zeta_j|\boldsymbol{\vartheta}) \right\} p(\boldsymbol{\vartheta}). \tag{10}$$

Using common parameters at the population level allows us to pool information about typical patterns of demand, and the product-level parameters capture the heterogeneity exhibited by the particular products (see Gelman and Hill 2006, for a general discussion).

4.2. Use of the Weibull Curve

Like the forecasting model in (Yelland 2010), the model in this paper incorporates an explicit representation of the product's life cycle. The representation used in this model is derived from the *Weibull distribution*, following a precedent established by Sharif and Islam (1980) and Moe and Fader (2002), who use the Weibull in the analysis of innovation diffusion and new product adproduct, respectively.[7]

In Moe and Fader's (2002) model, use of a Weibull curve to describe the time to first purchase of a new product has theoretical appeal, deriving from the Weibull distribution's origins in the analysis of events that occur after a period of random duration (McCool 2012). Observe, however, that the Weibull distribution plays no explicit rôle in the data generating process described in Section 3, and given the intricacies of the simulation, a foundational argument the use of the Weibull curve to describe it is difficult to make. Thus the application of the Weibull curve here follows Sharif and Islam's (1980)'s more pragmatic approach—as both they (and in fact Moe and Fader) observe, the Weibull curve brings an appealing combination of parsimony and flexibility to the description of life cycles. By way of justification for this pragmatic approach, an informal indication of how well the Weibull curve—with only two parameters—captures the trend in the simulated product demands can be gauged from Figure 1, where (appropriately scaled) Weibull curves have been fitted to the sample series.

4.3. Determinants of Product Demand

Returning to the discussion of section 4.1, the first step in making concrete the general model of equation (6), is to specify the conditional distribution $p(y_{jt}|\boldsymbol{\theta})$ of unit demand for product j in period t. Since product demand in any time period is a discrete, non-negative quantity, it is natural to represent it as a Poisson distribution with a time-varying mean. The latter is the product of three random quantities: 1) a product-specific quantity, γ_j, that depends on the potential market for the product, 2) a discrete-time stochastic process that captures the product's life cycle, λ_{jt}, and 3) a factor ς_{jt} associated with seasonal effects. Thus:

$$y_{jt} \sim \mathrm{Pois}(\gamma_j \lambda_{jt} \varsigma_{jt}). \tag{11}$$

[7] For a comprehensive survey of analytical models of product life cycles, see Mahajan et al. (2000).

4.4. Scale Factor

Filling out the specification in equation (11), we must provide a prior for the scale factor γ_j. Since this quantity is necessarily positive, its prior is a left-truncated normal distribution whose location and scale parameters are shared with other products; these shared (hyper-) parameters are themselves given non-informative priors:

$$\gamma_j \sim \mathrm{N}_{[0,\infty)}(\mu_\gamma, \sigma_\gamma^2), \quad p(\mu_\gamma) \propto 1, \quad p(\sigma_\gamma) \propto \mathcal{I}(\sigma_\gamma > 0). \tag{12}$$

4.5. Life Cycle Curve

Following the plan set out in section 4.3, the quantity λ_{jt} in equation (11), which traces the evolution of demand over the life cycle of an product, is determined by a suitably parameterized Weibull probability density function (*PDF*) at t:

$$\lambda_{jt} = \mathrm{Weib}(t|\alpha_j, \delta_j). \tag{13}$$

In the conventional parameterization of the Weibull curve, the value of the Weibull PDF at t is equal to $(\eta/k)(t/k)^{\eta-1}e^{-(t/k)^\eta}$, where η and k are (positive) parameters of the distribution. To help the convergence of the MCMC simulator described in Section 5, and as an aid to interpretability, we use an alternate parameterization of the Weibull in equation (13), indexed by α_j and δ_j, which are respectively the 20^{th} percentile of the distribution and the difference between its 95^{th} and 20^{th} percentiles.[8] A little algebraic manipulation yields conventional Weibull parameters η_j and k_j corresponding to α_j and δ_j, so that:

$$\mathrm{Weib}(t|\alpha_j, \delta_j) = \frac{\eta_j}{k_j}\left(\frac{t}{k_j}\right)^{\eta_j - 1} e^{-(t/k_j)^{\eta_j}},$$

$$\text{where } \eta_j = \frac{2.6}{\log(\alpha_j + \delta_j) - \log(\alpha_j)}\ , \ k_j = \frac{\alpha_j}{0.22^{1/\eta_j}}\ . \tag{14}$$

Completing the specification of λ_{jt} given by equations (13) and (14) requires that we provide prior distributions for the part-specific parameters α_j and δ_j. As with γ_j, the priors for both of these parameters are hierarchical, incorporating information garnered from the demand for other products. Treating the case of α_j in detail (the treatment of δ_j is analogous): Since α_j must be positive in equation (14), it is drawn (like γ_j) from a normal distribution, truncated on the left at 0.[9] The location and scale parameters, μ_α and σ_α, resp. of this truncated normal distribution are common to all parts, and have non-informative priors. In symbols:

$$\alpha_j \sim \mathrm{N}_{[0,\infty)}(\mu_\alpha, \sigma_\alpha^2), \quad p(\mu_\alpha) \propto 1, \quad p(\sigma_\alpha) \propto \mathcal{I}(\sigma_\alpha > 0). \tag{15}$$

4.6. Seasonal Effects

The representation of seasonal effects in the model mirrors that used in the simulation (section 3).[10] Thus demand is modulated by an product-specific multiple ω_j of the indicator variable $\mathtt{S}_t$ from equation (2). To complete the specification, ω_j is drawn from a normal distribution with pooled hyperparameters:

$$\varsigma_{jt} = 1 + \omega_j \mathtt{S}_t, \quad \omega_j \sim \mathrm{N}(\mu_\omega, \sigma_\omega^2), \quad p(\mu_\omega) \propto 1, \quad p(\sigma_\omega) \propto \mathcal{I}(\sigma_\omega > 0). \tag{16}$$

---

[8]We use the difference between the 20^{th} and 95^{th} percentiles rather than the 95^{th} percentile itself because α_j and δ_j might reasonably be considered *a priori* independent, making for easier specification of the model prior.

[9]Strictly speaking, truncation on the left should be at a point slightly greater than 0, but the technical elision is of no practical consequence.

[10]In this respect, we depart from the precept set out in the introduction requiring us to disassociate the simulation and forecasting model. However, since we do not regard seasonal effects as a defining characteristic of demand for fashion goods (to which life cycle effects are most pivotal), and since alternative approaches to seasonal modeling would be unnecessarily cumbersome here, we consider such a lapse justified.

Product demand

$$y_{jt} \sim \mathrm{Pois}(\gamma_j \lambda_{jt} \varsigma_{jt})$$

Life cycle curve

$$\lambda_{jt} = \mathrm{Weib}(t|\alpha_j, \delta_j), \quad \alpha_j \sim \mathrm{N}_{[0,\infty)}(\mu_\alpha, \sigma_\alpha^2), \quad \delta_j \sim \mathrm{N}_{[0,\infty)}(\mu_\delta, \sigma_\delta^2)$$

Scale

$$\gamma_j \sim \mathrm{N}_{[0,\infty)}(\mu_\gamma, \sigma_\gamma^2)$$

Seasonal effects

$$\varsigma_{jt} = 1 + \omega_j \mathrm{s}_t, \quad \omega_j \sim \mathrm{N}(\mu_\omega, \sigma_\omega^2)$$

Figure 2: Model summary: Unless otherwise stated, location parameters of the form $\mu_\circ$ and scale parameters $\sigma_\circ$ have non-informative priors $p(\mu_\circ) \propto 1$ and $p(\sigma_\circ) \propto \mathcal{I}(\sigma_\circ > 0)$, respectively. Indexes j and t range over products and periods, resp.

5. Estimation

This section discusses the procedure used to approximate the posterior distributions of the model parameters, and summarizes the approximate posteriors produced with product demands simulated as described in Section 3.

5.1. Metropolized Gibbs Sampler

As is commonly the case in modern applied Bayesian statistics, estimation of posterior distributions for the parameters of the model described in the previous section is carried out using a Markov chain Monte Carlo simulator—specifically, an adaptation of the *Gibbs sampler* known variously as the *Metropolized-* or the *Metropolis-within-*Gibbs sampler (Robert and Casella 2005, p. 392). Schematic descriptions of the Gibbs sampling scheme now abound in the literature—see (Gilks et al. 1996, chp. 1), for example; briefly, beginning with starting values for the parameters of the model, such a simulator constructs a Markov chain whose states converge to a dependent sample from the joint posterior of those parameters. Each transition in this Markov chain involves drawing a new value of one of the parameters from its posterior distribution conditional on the current value of the other parameters and the observed data. The *Metropolized* Gibbs sampler—first proposed by Müller (1991)—adopts the same step-by-step sampling scheme as the ordinary Gibbs sampler, but *Metropolis-Hastings sampling* (also discussed by Gilks et al.) is used to sample approximately from the conditional distributions (by accepting or rejecting so-called *proposed values*) for those steps where sampling directly from the conditional distribution is problematic. The individual steps of this particular sampler are described below. Many of the direct sampling steps rely on standard results concerning conjugate updating in Bayesian analysis, which may be found in texts such as (Bernardo and Smith 1994) or (Gelman et al. 2003). Where Metropolis-Hastings sampling is used, proposed values are generated using Gilks et al.'s (1995) *adaptive rejection Metropolis sampling (ARMS)* procedure, as implemented in the R package `dlm` (Petris 2010).

In the following, each step is introduced by the conditional distribution from which a sample is to be drawn. Variables of which the sampled quantity is conditionally independent are omitted from the conditioning set. In the interests of brevity, draws are specified only for α_j and σ_α; samples for δ_j and σ_δ are generated in an analogous manner. We abbreviate $\boldsymbol{y}_j = (y_{j1}, \ldots, y_{jT_j})$.

$\gamma_j \mid \boldsymbol{y}_j, \alpha_j, \delta_j, \mu_\gamma, \sigma_\gamma, \omega_j$

The kernel of the full conditional distribution is given by the expression:

$$\left[\prod_{t=1}^{T_j} \mathrm{Pois}(y_{jt}|\gamma_j \lambda_{jt} \varsigma_{jt})\right] \times \mathrm{N}_{[0,\infty)}(\gamma_j|\mu_\gamma, \sigma_\gamma^2),$$

where λ_{jt} and ς_{jt} are the quantities determined by α_j, δ_j and ω_j in equation (13) and equation (16) resp. Sampling is carried out using the ARMS procedure.

$\alpha_j \mid \boldsymbol{y}_j, \gamma_j, \delta_j, \mu_\alpha, \sigma_\alpha, \omega_j$

The full conditional is proportional to the expression:

$$\left[\prod_{t=1}^{T_j} \mathrm{Pois}(y_{jt}|\gamma_j \lambda_{jt} \varsigma_{jt})\right] \times \mathrm{N}_{[0,\infty)}(\alpha_j|\mu_\alpha, \sigma_\alpha^2).$$

This is also sampled using ARMS.

$\mu_\alpha \mid \alpha_1, \ldots, \alpha_J, \sigma_\alpha$

Sampling is carried out using a device due to Griffiths (2004): Specifically, for $l \in 1, \ldots, J$, let:

$$\tilde{\alpha}_l = \mu_\alpha + \sigma_\alpha \, \Phi^{-1}\left[\frac{\Phi\left(\frac{\alpha_l - \mu_\alpha}{\sigma_\alpha}\right) - \Phi\left(\frac{-\mu_\alpha}{\sigma_\alpha}\right)}{1 - \Phi\left(\frac{-\mu_\alpha}{\sigma_\alpha}\right)}\right], \tag{17}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

Then as Griffiths demonstrates, supposing that $\tilde{\alpha}_l \sim \mathrm{N}(\mu_\alpha, \sigma_\alpha^2)$ and drawing from the conditional distribution $\mu_\alpha \mid \tilde{\alpha}_1, \ldots, \tilde{\alpha}_J, \sigma_\alpha$ (a straightforward application of semi-conjugate updating with a non-informative prior) is equivalent to drawing from $\mu_\alpha \mid \alpha_1, \ldots, \alpha_J, \sigma_\alpha$ given that $\alpha_l \sim \mathrm{N}_{[0,\infty)}(\mu_\alpha, \sigma_\alpha^2)$.

$\sigma_\alpha \mid \alpha_1, \ldots, \alpha_J, \mu_\alpha$

Again, using Griffiths's device, draw from $\sigma_\alpha \mid \tilde{\alpha}_1, \ldots, \tilde{\alpha}_J$, given that $\tilde{\alpha}_j \sim \mathrm{N}(\mu_\alpha, \sigma_\alpha^2)$, where $\tilde{\alpha}_j$ is defined in equation (17).

$\omega_j \mid \boldsymbol{y}_j, \gamma_j, \alpha_j, \delta_j, \mu_\omega, \sigma_\omega$

Another ARMS step, with kernel:

$$\left[\prod_{t=1}^{T_j} \mathrm{Pois}(y_{jt}|\gamma_j \lambda_{jt}[1 + \omega_j \mathtt{S}_t])\right] \times \mathrm{N}(\omega_j|\mu_\omega, {\sigma_\omega}^2).$$

5.2. Posterior Estimates and Diagnostics

Figure 3 (inspired by similar figures on e.g. p. 351 of Gelman and Hill 2006) illustrates the results of estimating the model with the simulated product demands. Here the Gibbs sampler was run in a single

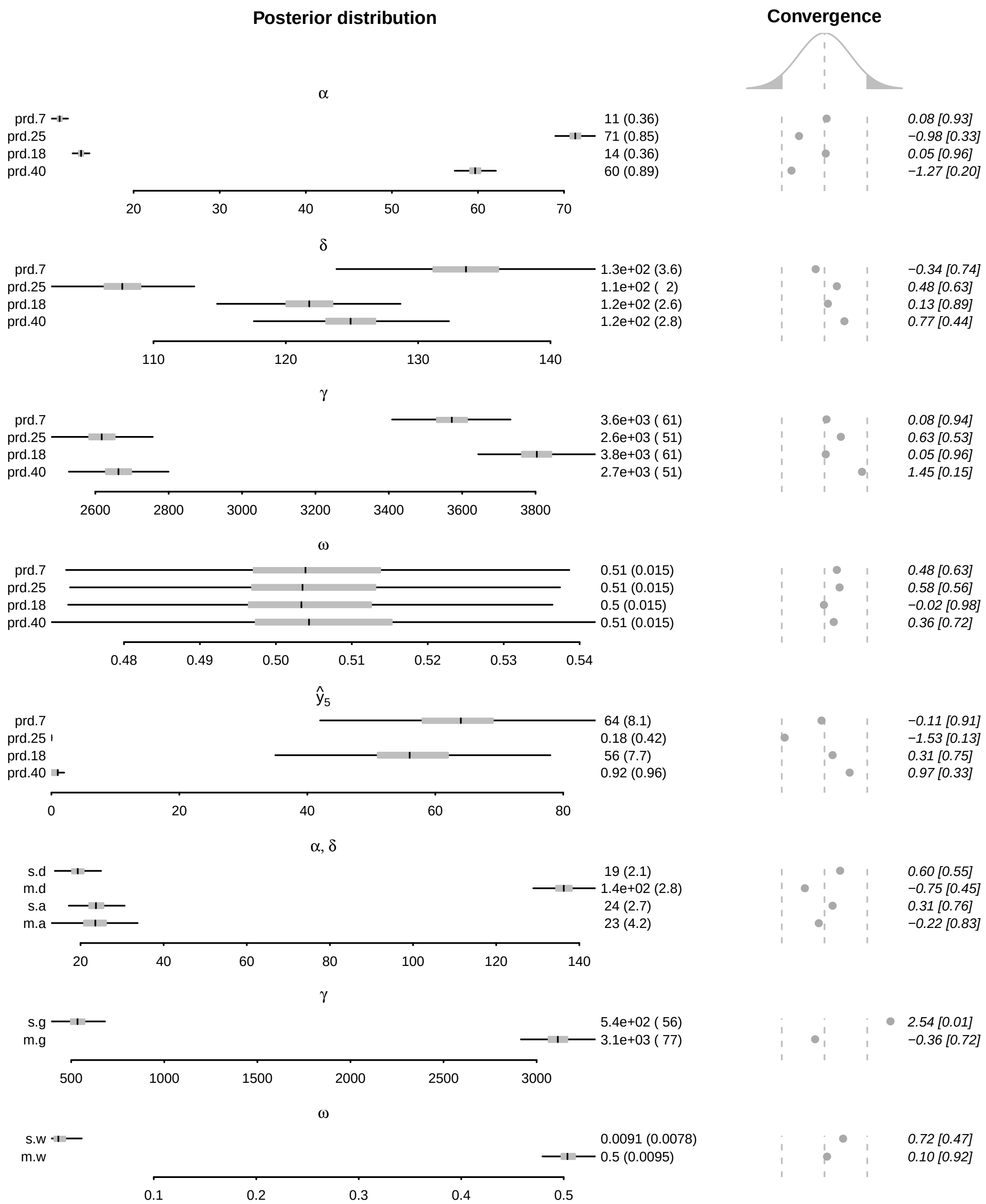

Figure 3: Posterior distributions and convergence diagnostics for model fit to full sample of simulated demand series: On the left are interquartile ranges and 1.5×interquartile ranges of quantities associated with selected products (`prd.-`) and hyperparameters, together with the mean and standard deviation of the corresponding distribution. On the right are values of Geweke's convergence statistic; all but one of the values lie between the 5^{th} and 95^{th} percentiles of the standard normal distribution (statistic values and p-values appear on the far right), so convergence is reasonably assured.

chain for 6,000 iterations, with samples from the first 1,500 discarded; no thinning of the remaining samples was performed.

On the left of the Figure are displayed posterior distributions for the quantities associated with a random selection of products ("prd."- 7, 25, 18 and 40), as well as population-level parameters. Parameters α_j, δ_j and γ_j are summarized for the selected products, and the fitted value of y_{jt} in the fifth period is displayed as "y.5". Also shown are the population-level location parameters μ_α, μ_δ and μ_γ (labeled respectively "m.a", "m.d", "m.g" and "m.w"), as well as scale parameters σ_α, σ_δ, σ_γ and σ_ω (similarly labeled "s.a", etc.). Each posterior distribution is summarized graphically by a condensed box-and-whisker plot,[11] with the distribution's mean and (parenthesized) standard deviation given numerically.

The right-hand side of the Figure plots Geweke's (1992) convergence diagnostic for each of the quantities in question. Derived from a comparison of the first and last segments of the Markov chain associated with a quantity, Geweke's statistic z has an asymptotically standard normal distribution if the chain is stationary (i.e. convergence has occurred). On the diagram, values of z are plotted on the 5^{th} and 95^{th} percentiles of the standard normal distribution. The plots indicate that convergence has been achieved—note that with 28 quantities displayed, we would expect between 2 and 3 values of z to fall outside of the percentile bounds even with convergence.

6. Testing Forecast Performance

The test the efficacy of the model described in the foregoing, we used it to forecast demand for the sample of product demand series simulated using the framework described in section 3. As Armstrong (2001) points out, an objective assessment of new forecasting method requires a *benchmark* method against which to compare its performance. In this exercise, our benchmark method was the `ets` function found in the "`forecast`" R package of Hyndman and Khandakar (2008). The `ets` function embodies the forecasting framework developed by Hyndman et al. (2002) which—as its authors point out—has been proven to deliver consistently superior forecast performance when applied to a wide variety of time series. Furthermore, the `ets` function is capable of producing forecasts in a largely automatic fashion (though its behavior may be affected by parameter settings). Thus while the `ets` function—unlike the hierarchical Bayesian method described in this paper—is not tailored specifically to the forecasting of fashion goods demand, it promises in its rôle as a benchmark method to furnish reasonable performance with minimal effort.

Our testing procedure was as follows:

1. Reserve 25 out of the 50 simulated demand series from section 3 to calibrate the Bayesian forecasting model.

2. Truncate all of the remaining 25 series after k periods, where k takes successive values from 4 to 92 in increments of 8 (i.e. 4, 12, 20, etc.).

3. Using the hierarchical Bayes model and the benchmark `ets` function in turn, for each value of k, and each truncated series $\boldsymbol{s}_i = (s_{i1}, \ldots, s_{ik})$, produce two forecasts $\hat{s}^{\texttt{hb}}_{i,k+h}$ and $\hat{s}^{\texttt{ets}}_{i,k+h}$ of demand h periods after the point of truncation, where h ranges from 1 to $100 - k$ (since all of the simulated series are of length 100 periods).

   Forecasting with the Bayesian model involves subjecting both the 25 reserved demand series as well the 25 truncated series to the procedure set out in Section 4.1, using the MCMC simulator described in Section 5.1 to produce a sample from the posterior predictive distribution for $s_{i,k+h}$, and taking the median of this sample as the point forecast.

   The `ets` function did not use the calibration series, but we found that to produce reasonable forecasts with the function at longer horizons (i.e. h in excess of 20 or so), it was necessary to provide some hints to its automatic model selection algorithm, constraining it to look for models with *damped, additive trends* (Gardner and McKenzie 1989, Hyndman et al. 2002).

[11] "Boxes" delimit the interquartile range of the distributions, and "whiskers" extend 1.5 times the interquartile range from the ends of the boxes—see Tukey (1977) for further details.

4. Subtracting the predicted value for $s_{i,k+h}$ from its actual value in the untruncated series produces two forecast errors. To derive *scale-free* expressions of forecast error (in the sense of Hyndman and Koehler 2006), these were expressed as a percentage of the mean value of the untruncated series,[12] yielding quantities $\epsilon^{\mathtt{hb}}_{i,k+h}$ and $\epsilon^{\mathtt{ets}}_{i,k+h}$.

5. Finally, overall performance metrics for each method for each value of k and h were calculated by taking means of the absolute values of the scale free errors across all of the truncated series—thus $\bar{\epsilon}^{\mathtt{hb}}_{k+h} = \sum_{i=1}^{25} \epsilon^{\mathtt{hb}}_{i,k+h}$ and $\bar{\epsilon}^{\mathtt{ets}}_{k+h} = \sum_{i=1}^{25} \epsilon^{\mathtt{ets}}_{i,k+h}$.

Figure 4 plots the forecast metrics calculated in step 5 for all values of k (the number of periods observed before forecasting). It is evident that in this application, the Bayesian model proposed here offers substantially better forecast performance, particularly early in the product lifecycle, where the calibration data comprised by the reserved demand series is especially advantageous.[13]

7. Conclusions

We set out to establish the effectiveness of a Bayesian forecasting model based on a general formulation of the product life cycle in application to fashion clothing goods. The last section demonstrates that the forecasting performance of the proposed model certainly compares favorably with that of a benchmark method when applied to simulated fashion product demand series. Of course, a caveat attaches to this demonstration, in that our data is simulated; though we have been careful to derive the simulation from a generally-accepted model of consumer behavior with regard to fashion products, this is provides only a provisional assurance that the simulated series match the actual series encountered by a particular retailer or manufacturer.[14] Nonetheless, the archetypical features of the simulated data implies that the forecasting model is probably at least partially applicable in any given situation in practice, though it may require some adaptation.

Certainly, as suggested in section 1, many elaborations of both the simulation and the forecasting model—to accommodate more realistic seasonal effects, for example—are reasonably straightforward. (Yelland 2010), for example, illustrates the incorporation of a latent autoregressive process to account for serially-correlated departures of demand from a strict life cycle model. And as also shown in (Yelland 2010), a more elaborate hierarchical Bayesian model can be used to forecast demand for ancillary items, such as configurations or parts, that are correlated with product demand. And the related system documented in (Yelland et al. 2010) uses substantive Bayesian priors to predict life cycle demands in a situation where—unlike the forecasting test of section 6—comparable products are unavailable for calibration. The Bayesian framework may also be employed to address another problem Bruce and Daly (2007) observe is often encountered in supply chain management for fashion retailing, viz., the availability of timely sales data that is only approximate and/or provisional. With a Bayesian model, revised forecasts can be produced by treating early sales data as *soft evidence*, known only with uncertainty—see e.g. (Pearl 1988, Valtorta et al. 2002, Pan et al. 2006) for details of methods for updating Bayesian models with soft evidence.

With regard to applying our work in an industry setting, as we suggest in section 1, recent developments in the fashion industry are highly favorable to a life cycle based Bayesian model such as the one described here: First, as Hines and Bruce (2007) indicate, increasing globalization of the industry—driven by the dismantling of tarrif barriers, improvements in transportation and communication technology, competitive pressures militating in favor of low-cost off-shore manufacturing and the rise of geographically-dispersed

[12]This is a modification of the more common *mean absolute percentage error* (MAPE) metric; since the values in the forecast series vary significantly over lifecycle of the corresponding product, the MAPE is apt to distort performance assessment by unduly inflating small absolute errors at either end of the lifecycle.

[13]Note that the comparatively poor performance of the `ets` function here should not be construed as general criticism, but rather as affirmation of the contention that the proposed model is far better suited to the type of series encountered in this context than is a general-purpose forecasting tool.

[14]In this respect, the situation here is the obverse of that in (Yelland 2010), where efficacy for a particular manufacturer's demand series was demonstrated, but more general applicability remained something of an open question.

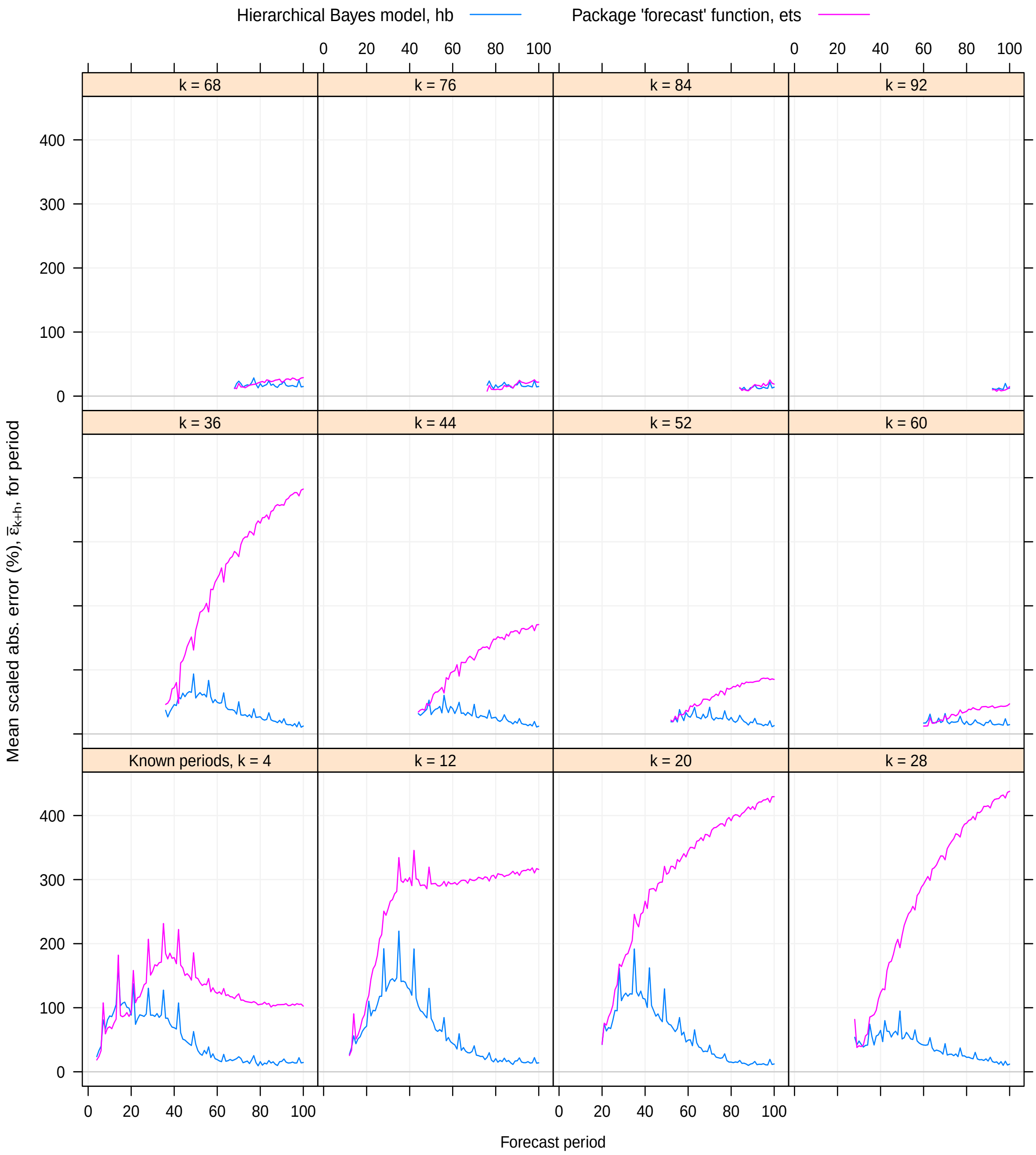

Figure 4: Forecast performance of the model and benchmark.

design and manufacturing hubs in Europe and the Far East—has made relatively long lead times a reality in many sectors of the industry. Effectively addressing such lead times necessitates relatively long-term forecasts, in which product life cycles are frequently a significant factor. Second, as competitive pressures accelerate the trend towards "fast fashion" and "quick response" (Kang 1999) even in low-cost retail lines, updates to forecasts prior to and during the selling season are becoming the norm;[15] Bayesian forecasting of the sort demonstrated here is perfectly suited to producing such updates. We would also be interested in establishing the broader viability of the use of agent-based simulation models as a means of calibrating the efficacy of different forecasting models *a priori*. Agent-based simulation is already a proven approach to operations planning for supply chains (see (Fox et al. 2000), for example), but to our knowledge, the use of agent-based simulation to compare forecasting models is an area of research still to be explored. With its continuing protean changes in sourcing and retailing arrangements, the fashion industry constitutes an ideal venue for such work.

References

N. Agrawal and S. A. Smith. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics*, 43(6):839–861, 1996.

A. Ansari, K. Jedidi, and S. Jagpal. A hierarchical Bayesian methodology for treating heterogeneity in structural equation models. *Marketing Science*, 19(4):328–347, 2000.

J. M. Armstrong. Standards and practices for forecasting. In J. M. Armstrong, editor, *Principles of Forecasting: A Handbook for Researchers and Practitioners*, pages 171–192. Kluwer, Norwell, MA, 2001.

K. S. Azoury. Bayesian solution to dynamic inventory models under unknown demand distribution. *Management Science*, 31:1150–1160, 1985.

K. S. Azoury and B. S. Miller. A comparison of the optimal levels of Bayesian and non-Bayesian inventory models. *Management Science*, 30:993–1003, 1984.

F. Bass. A new product growth model for consumer durables. *Management Science*, 15:215–227, 1969.

J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley, Hoboken, NJ, 1994.

P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2nd edition, 2002.

M. Bruce and L. Daly. Challenges of fashion buying and merchandising. In T. Hines and M. Bruce, editors, *Fashion Marketing: Contemporary Issues*, chapter 3, pages 54–69. Elsevier, 2nd edition, 2007.

F. Caro and J. Gallien. Dynamic assortment with demand learning for seasonal consumer goods. *Management Science*, 53(2):276–292, 2007.

T. M. Choi, D. Li, and H. M. Yan. Optimal two-stage ordering policy with Bayesian information updating. *Journal of the Operational Research Society*, 54(8):846–859, 2003.

T. M. Choi, D. Li, and H. M. Yan. Quick response policy with Bayesian information updates. *European Journal of Operational Research*, 170(3):788–808, 2006.

M. D. Cohen, J. G. March, and J. P. Olsen. A garbage can model of organizational choice. *Administrative Science Quarterly*, 17:1–25, 1972.

S. R. Cook, A. Gelman, and D. B. Rubin. Validation of software for Bayesian models using posteior quantiles. *Journal of Computational and Graphical Statistics*, 15(3):675–692, 2006.

S. A. Delre, W. Jager, T. H. A. Bijmolt, and M. A. Janssen. Targeting and timing promotional activities: An agent-based model for the takeoff of new products. *Journal of Business Research*, 60(8):826–835, 2007.

X. Dong, P. Manchanda, and P. K. Chintagunta. Quantifying the benefits of individual-level targeting in the presence of firm strategic behavior. *Journal of Marketing Research*, 9(3):301–337, 2009.

[15] Hines and Bruce (2007) point out that "fast fashion" retailers such as Zara and Primark publish revised forecasts on a weekly basis

K. L. Donohue. Efficient supply contracts for fashion goods with forecast updating and two production modes. *Management Science*, 46(11):1397–1411, 2000.

A. Dvoretzky, J. Kiefer, and J. Wolfowitz. The inventory problem: Case of known distribution of demand. *Econometrica*, 20:187–222, 1952.

G. D. Eppen and A. V. Iyer. Improved fashion buying with Bayesian updating. *Operations Research*, 45(6): 805–819, 1997.

M. S. Fox, M. Barbuceanu, and R. Teigen. Agent-oriented supply-chain management. *Intl. Journal of Flexible Manufacturing Systems*, 12:165–188, 2000.

E. S. Gardner and E. McKenzie. Seasonal exponential smoothing with damped trends. *Management Science*, 35(3):372–376, March 1989.

A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3): 515–533, 2006.

A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, New York, 2006.

A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis.* Chapman & Hall/CRC Press, Boca Raton, FL, 2nd edition, 2003.

J. Geweke. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 4, pages 169–193. Clarendon Press, Wotton-under-Edge, Gloucestershire, UK, 1992.

W. R. Gilks, N. G. Best, and K. K. C. Tan. Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, 44:455–472, 1995. (Corr: Neal, R.M. 1997 v.46 pp.541-542).

W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice.* Chapman & Hall, Boca Raton, FL, 1996.

J. Goldenberg, S. Han, D. R. Lehmann, and J. W. Hong. The role of hubs in the adoption process. *Journal of Marketing*, 73(2):1–13, 2009.

W. Griffiths. A Gibbs sampler for the parameters of a truncated multivariate normal distribution. In R. Becker and S. Hum, editors, *Contemporary Issues in Economics and Econometrics: Theory and Application*, pages 75–91. Edward Elgar, 2004.

H. Gurnani and C.S. Tang. Note: Optimal ordering decisions with uncertain cost and demand forecast updating. *Management Science*, 45(10):1456–1462, 1999.

A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge, UK, 1989.

T. Hines and M. Bruce. *Fashion Marketing.* Elsevier, Oxford, 2007.

N. A. Hunter and P. Valentino. Quick response—ten years later. *International Journal of Clothing Science and Technology*, 7:30–40, 1995.

R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3):1–22, July 2008. URL http://www.jstatsoft.org/v27/i03.

R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679–688, 2006.

R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, 2002.

D. Iglehart. The dynamic inventory problem with unknown demand distribution. *Management Science*, 10: 429–440, 1964.

A. Iyer and M. E. Bergen. Quick response in manufacturer-retailer channels. *Management Science*, 43(4): 559–570, 1997.

K.-Y. Kang. *Development of an Assortment-Planning Model for Fashion-Sensitive Products*. PhD thesis, Virginia Polytechnic Institute and State University, 1999.

E. Kiesling, M. Gunther, C. Stummer, and L .M. Wakolbinger. An agent-based simulation model for the market diffusion of a second generation biofuel. In *Winter Simulation Conference*, 2009.

K. Krishnamoorthy. *Handbook of Statistical Distributions with Applications*. Statistics: A Series of Textbooks and Monographs. Chapman and Hall/CRC, 2006.

V. Mahajan, E. Muller, and Y. Wind, editors. *New-Product Diffusion Models*. Kluwer, Norwell, MA, 2000.

J. I. McCool. *Using the Weibull Distribution: Reliability, Modeling, and Inference*. Wiley, 2012.

B. Miller. Scarf's state reduction method, flexibility and a dependent demand inventory model. *Operations Research*, 34(1):83–90, 1986.

W. W. Moe and P. S. Fader. Using advance purchase orders to forecast new product sales. *Marketing Science*, 21(3):347–364, Summer 2002.

P. Müller. A generic approach to posterior integration and Gibbs sampling. Technical Report 91-09, Department of Statistics, Purdue University, 1991.

G. R Murray, Jr. and E. A. Silver. A Bayesian analysis of the style goods inventory problem. *Management Science*, 12(11):785–797, 1966.

S. Narayanan and P. Manchanda. Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Science*, 28(3):424–441, 2009.

M. J. North, C. M. Macal, J. St. Aubin, P. Thimmapuram, M. Bragen, and J. Hahn. Multi-scale agent-based consumer market modeling. *Complexity*, 15(5):37–47, 2010.

R. Pan, Y. Peng, and Z. Ding. Belief update in bayesian networks using uncertain evidence. In *Intl. Conf. on Tools with Artificial Intelligence, ICTAI*, pages 441–444, 2006.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1988.

G. Petris. dlm: Bayesian and likelihood analysis of dynamic linear models, 2010. URL http://cran.r-project.org/web/packages/dlm/index.html.

W. Rand and R. T. Rust. Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing*, 28(3):181–183, September 2011.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, 2nd edition, 2005.

P. E. Rossi, R. E. McCulloch, and G. M. Allenby. The value of purchase history data in target marketing. *Marketing Science*, 15(4):321–340, 1996.

P. E. Rossi, G. M. Allenby, and R. E. McCulloch. *Bayesian Statistics and Marketing*. Wiley, 2005.

H. Scarf. Bayes solutions of the statistical inventory problem. *Annals of Mathematical Statistics*, 30(2): 490–508, 1959.

T. A. Schenk, G. Loffler, and J. Rauh. Agent-based simulation of consumer behavior in grocery shopping on a regional level. *Journal of Business Research*, 60(8):894–903, 2007.

M. Schwoon. Simulating the adoption of fuel cell vehicles. *Journal of Evolutionary Economics*, 16(4): 435–472, 2006.

M. N. Sharif and M. N. Islam. The Weibull distribution as a general model for forecasting technological change. *Technological Forecasting and Social Change*, 18:247–256, 1980.

G. B. Sproles. Analyzing fashion life cycles: Principles and perspectives. *Journal of Marketing*, 45:116–124, 1981.

J.W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

M. Valtorta, K. Young-Gyun, and J. Vomlel. Soft evidential update for probabilistic multiagent systems. *Intl. Journal of Approximate Reasoning*, 29:71–106, 2002.

W. Walsh and M. Wellman. Modeling supply chains formation in multi-agent systems. In *IJCAI-99 Work-*

shop Agent-Mediated Electronic Commerce, 1999.

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:409–410, 1998.

D. S. Wu, K. G. Kempf, M. O. Atan, B. Aytac, S. A. Shirodkar, and A. Mishra. Improving new product forecasting at Intel corporation. *Interfaces*, 40(5):385–396, 2010.

P. M. Yelland. Bayesian forecasting of parts demand. *International Journal of Forecasting*, 26(2):374–396, April–June 2010.

P. M. Yelland, S. Kim, and R. Stratulate. A Bayesian model for sales forecasting at Sun Microsystems. *Interfaces*, 2010.

Appendices

Appendix A. Notation

Notation	Meaning
	Indexes
$i \in \{1, \ldots, N\}$	Indexes ranging over agents.
$t \in \{1, \ldots, T\}$	Index ranging over planning periods.
	Simulation
$\texttt{Neighbors}_i$	The neighbors of agent i.
$\texttt{Market}$	The potential market for a given product.
$\texttt{Purch}_t$	Set of agents who have purchased the product by time t.
$\texttt{pot}_{it}$	Agent i has a non-zero probability of purchase in period t.
$\texttt{purch-ex}_{it}, \texttt{purch-end}_{it}$	An ex-, resp. endogenously-motivated purchase by agent i in period t.
$\texttt{purch}_{it}$	A purchase (endo- or exogenously motivated) by agent i in period t.
$\texttt{S}_t$	"Saturday" indicator.
p_i, q_i	Parameter governing resp. ex-, endogenously-motivated purchases by agent i.
$\texttt{u-exo}_{it}, \texttt{u-endo}_{it}$	Random variable simulating ex-, endogenous period t purchase by agent i.
y_t	Simulated product demand during period t.
	Forecasting
$j \in \{1, \ldots, J\}$	Indexes ranging over similar products involved in forecast formulation.
T_j	Number of periods that demand for product j has been observed.
$\boldsymbol{y}_j = (y_{j1}, \ldots, y_{jT_j})$	Observed demand for product j.
$l \in \{1, \ldots, J\}$	Index of product for which forecast is to be produced.
h	Forecast horizon.
$y_{l\,T_l+h}$	Demand to be forecast.
$\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\vartheta}$	Vectors of unobserved quantities in a generic forecast model.
	Model (see also summary on p. 8)
$\mu_\theta, \sigma_\theta$	Parameters (usually mean and std. dev., resp.) of prior distribution for generic parameter θ.
γ_j	Scale factor associated with potential market for product j.
λ_{jt}	Stochastic process describing the life cycle of product j, delineating a Weibull curve.
α_j	20^{th} percentile of the Weibull distribution representing product j's life cycle.
δ_j	Difference between 95^{th} and 20^{th} percentiles of the Weibull distribution representing the product j's life cycle.
ς_{jt}	Demand variation produced by day of the week effect.

continued on next page

Notation	Meaning
	Testing
$\boldsymbol{s}_i$	Simulated demand series, truncated for testing.
k	The period after which simulated demand series was truncated.
$\hat{s}^{\texttt{ets}}_{i,k+h}, \hat{s}^{\texttt{hb}}_{i,k+h}$	Forecast value for simulated demand series $\boldsymbol{s}_i$ produced by benchmark `ets` function and proposed Bayesian model, resp.
$\epsilon^{\texttt{ets}}_{i,k+h}, \epsilon^{\texttt{hb}}_{i,k+h}$	Scale-free errors for benchmark and model forecasts.

Appendix B. Standard Probability Distributions

Below we list the probability distributions used in the paper, along with their standard parameterizations. For further details, see e.g. (Krishnamoorthy 2006).

Distribution	Description	Density/mass function
$\mathrm{Beta}(\alpha, \beta)$	Beta distribution with shape parameters α and β.	$\mathrm{Beta}(x\vert\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $x \in [0, 1]$
$\mathrm{Pois}(\lambda)$	Poisson distribution with parameter λ.	$\mathrm{Pois}(x\vert\lambda) = \frac{1}{x!}\lambda^x \exp(-\lambda)$, $x = 0, 1, \ldots$
$\mathrm{Weib}(\lambda, k)$	Weibull distribution with shape λ and scale k.	$\mathrm{Weib}(x\vert\lambda, k) = \lambda k^{-\lambda}\theta^{-1+\lambda} e^{-\left(\frac{\theta}{k}\right)^{\lambda}}$, $x \geq 0$
$\mathrm{N}(\mu, \sigma^2)$	Normal distribution with mean μ and standard deviation σ.	$\mathrm{N}(x\vert\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$
$\mathrm{N}_{[0,\infty)}(\mu, \sigma^2)$	The normal distribution $\mathrm{N}(\mu, \sigma^2)$, truncated on the left at 0.	$\mathrm{N}_{[0,\infty)}(x\vert\mu, \sigma^2) = 2\mathrm{N}(x\vert\mu,\sigma^2)$, $x \geq 0$
$\mathrm{Inv}\text{–}\chi^2(\nu)$	The inverse chi-squared distribution with ν degrees of freedom.	$\mathrm{Inv}\text{–}\chi^2(x\vert\nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{-(\nu/2+1)} \exp[-1/(2x)]$, $x > 0$