Bayesian Forecasting for Low Count Time Series Using State-Space Models: An Empirical Evaluation for Inventory Management

Phillip M. Yelland

Sun Microsystems Laboratories, 16 Network Circle, Menlo Park, CA 94025, U.S.A.

Abstract

Inventories of optional components in discrete manufacturing are often subject to so-called *low count* demand patterns. Quantities demanded from such inventories in any given period are sufficiently small that it may be unrealistic to forecast them with conventional models based on the normal distribution, and specialized models may be required. Fortunately, the statistical treatment of low count time series has been the focus of much recent research. This paper recounts an attempt to apply some of this research to forecasting demands for optional parts at Sun Microsystems, a manufacturer and vendor of network computer products. Specifically, we compare the forecast performance of three simple state-space models using demand data obtained from Sun's inventory management records. The models are estimated using Bayesian methods, producing forecasts in the form of full predictive distributions. The accuracy of these probabilistic forecasts is compared using techniques borrowed from the field of meteorology, allowing us to assess the suitability of the candidate models for this type of application.

Key words: Inventory management, low-count time series, Bayesian statistics, state-space models

1 Introduction

Figure 1 displays a number of time series comprising the units of a selection of manufacturing parts used over a 78-week period in the operations of Sun Microsystems Inc., a manufacturer and vendor of network computing products. The parts in the figure are a subset of a larger sample of some 100 optional parts (that is, parts whose inclusion in

Email address: phillip.yelland@sun.com (Phillip M. Yelland).

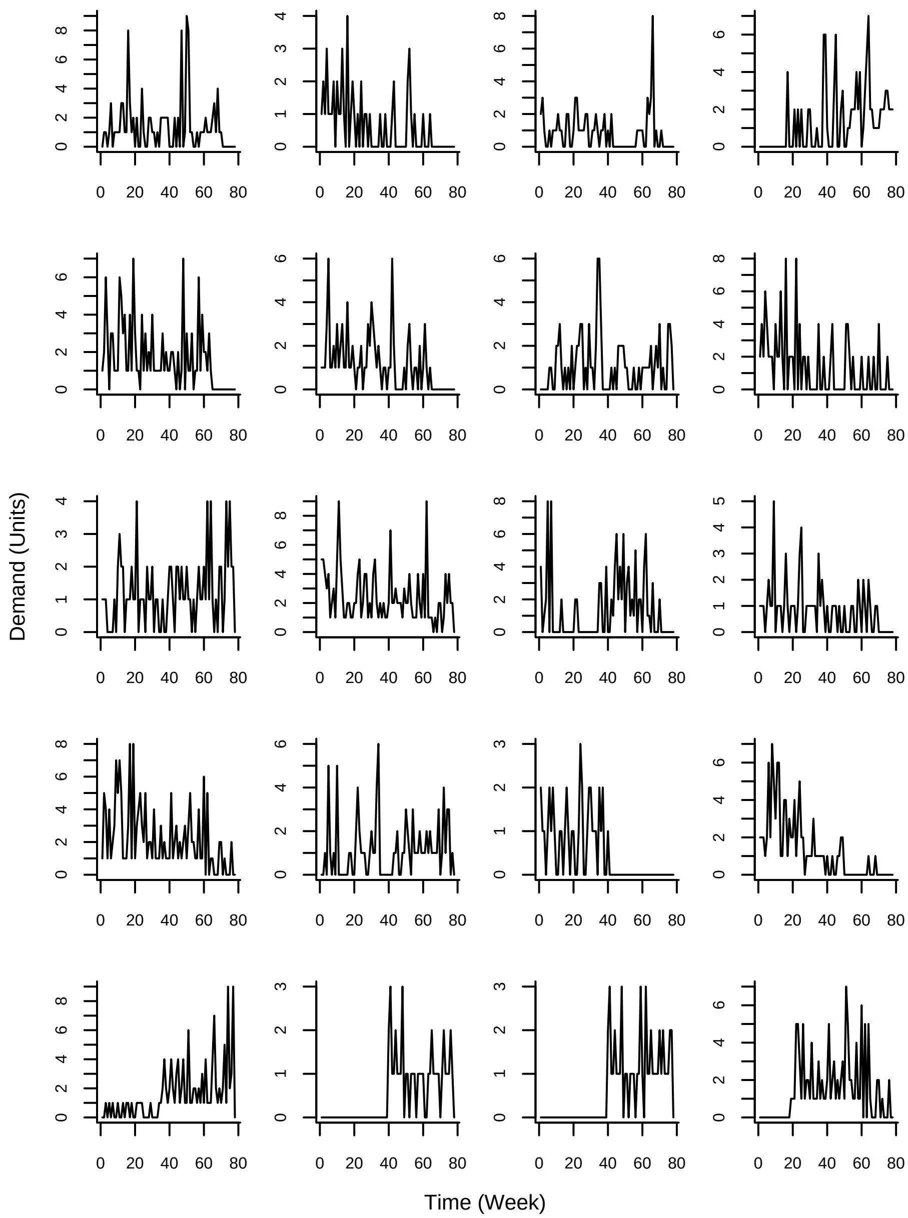

Figure 1. Sample inventory demands

a product depends upon configuration choices), drawn at random from Sun's inventory records.

Managing inventories for optional parts can be troublesome. In principle, with firm orders in hand for finished goods, materials requirements planning can be used to calculate parts demand over the short term through a straightforward bill-of-materials (*BOM*) "explosion" (Clement et al., 1995). In practice there are normally very many different types of such thinly-demanded components, so that the administrative overhead of entering them into the BOM and maintaining the correct BOM entries in the face of

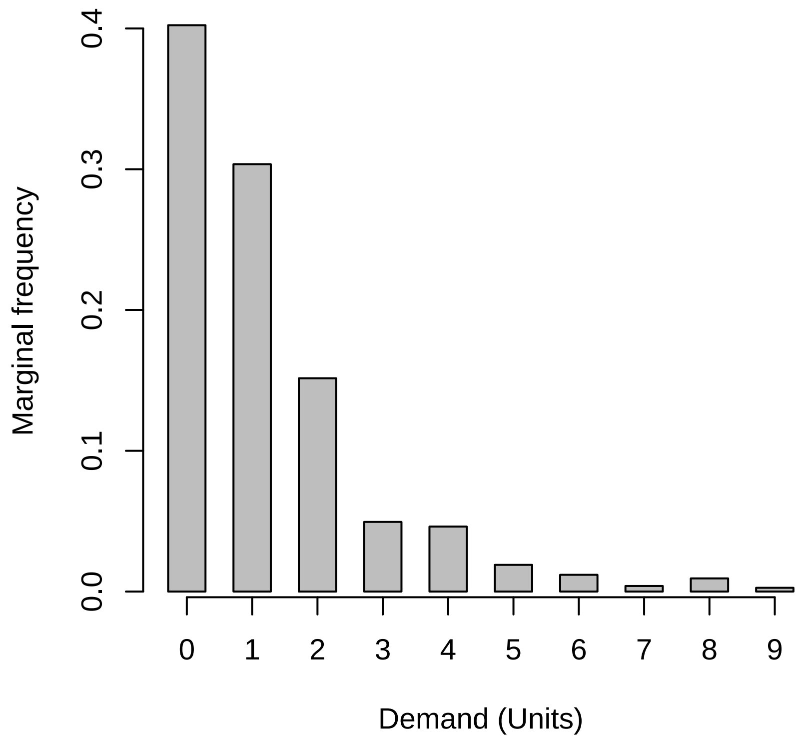

Figure 2. Marginal distribution of demand

product changes, changes in parts specification or supplier, etc., is often prohibitive. Planning at longer horizons could also be achieved by BOM explosion of finished good demand forecasts, but even if all the parts are actually in the BOM, their presence or absence in the final product depends on particular configuration choices, which must themselves be forecast. In many instances, therefore, it's often expedient to forecast demands for optional parts directly.

As Figure 1 illustrates, the time series in the sample are of a fairly idiosyncratic nature—an impression corroborated by Figure 2, which displays the marginal distribution of weekly demand across the entire sample. From the latter, it is clear that the bulk of the values in the series are positive integers between 0 and 4, with zero occurring quite frequently (in fact, weeks with 0 units of demand constitute approximately 40% of all the weeks in the sample). None of the parts experienced weekly demands in excess of nine units during the period of observation. McCabe and Martin (2005) refer to time series of this type as *low count series*, distinguished in that low counts are poorly approximated by that staple of mainstream forecasting models, the normal distribution. (In contrast, series comprising larger count values are approximated much more felicitously.)

2 Forecasting Low Count Time Series

Statistical modeling and prediction of low count time series has become the focus of much attention in recent years; surveys of this work may be found in (Cameron and Trivedi, 1998), (McKenzie, 2003) and (Winkelmann, 1997). Many authors in the field use Cox's (1981) taxonomy to distinguish two different types of low count model: In *observation-driven* models, dependence between values in a time-series is represented directly, usually by some form of autoregressive or moving average mechanism. A *parameter-driven* model, in contrast, uses an underlying latent process to induce dependence between observations. No model of either type has achieved the dominance that was long enjoyed by Box-Jenkins/ARIMA models in forecasting continuously-valued series. However, of the observation-driven models, the so-called *integer autoregressive* or *INAR* models described in (Al-Osh and Alzaid, 1990) and (McKenzie, 2003) have become increasingly popular. Correspondingly, most of the parameter-driven models use some form of *state space* formulation, wherein an unobserved state vector evolves according to a Markovian process, with observations conditional upon some function of the state vector. Thorough discussion of state space models may be found in (West and Harrison, 1997) and (Durbin and Koopman, 2001). In this paper, we concentrate on state-space models. While the potential of observation-driven models for inventory demand forecasting is not to be impugned, the state-space approach offers a number of advantages:

- As is demonstrated in (West and Harrison, 1997), the simple forms of state-space model in this paper may be thought of as direct formalizations of the familiar exponential smoothing methods frequently used in inventory planning. The forecasting performance of the models, therefore, has direct bearing on current practice.
- Simply by virtue of the fact that they are tied (possibly tenuously) to product lifecycles, sales of optional components are likely to be non-stationary. As (McKenzie, 2003) points out, however, the treatment of non-stationarity in observation-driven low count models is challenging. (Simple differencing, for example, radically alters the nature of the series, since it almost invariably results in negative values.) In contrast, non-stationarity is handled routinely by state-space models.
- Though the models examined in this paper have near-trivial state vectors, in general the components of the state vector in a state space model are amenable to intuitive explanation (as the "underlying level," "trend," "seasonal differences" and so on of the original series), which can greatly facilitate the communcation of forecasts to managers.
- On a technical level, in observation-driven models for low count series, the determinants of the correlation structure—being themselves observations—are *ipso facto* drawn from a restricted domain (viz. the positive integers). This usually necessitates the development of specialized mechanisms (such as the "binomial thinning" operator of the INAR models) to describe the correlation structure. No such considerations apply to state-space models, where elements of the state vector may be drawn from the entire real line, and the correlation structure of the series can be

expressed using conventional (usually linear) algebra.

Against these advantages must be weighed the computational effort required to estimate and extrapolate state-space models (especially the sort of state space models applied to low count series). With modern computing resources, of course, this is less of an issue than it was historically.

In keeping with much of the work involving state space models, we use Bayesian forecasting techniques as pioneered in the work of West and Harrison (1997). The Bayesian approach involves the provision of a prior for the initial value of the state vector, the computation of posterior state vector distributions for each observed time period, and full predictive distributions for series values in each of the forecast periods. Bayesian procedures offer a several benefits; most importantly for this application, the provision of full predictive distributions facilitates a comprehensive comparison of forecast performance, as we will see below. Bayesian estimation, of course, generally compounds the computational requirements of forecasting with state space models, but as (Yelland, 2004) reports, it is perfectly feasible with contemporary computers to use Bayesian state space-based forecasting as a routine part of manufacturing operations.

Parenthetically, since zero values feature so prominently in the sample series, they might be broadly characterized as expressions of *intermittent demand*, in the sense of (Boylan, 2005), for example. The main focus of forecasting for intermittent demand, however, concerns series with significant consecutive "runs" of zeroes, prompting the use of methods such as Croston's (1972) (Shenstone and Hyndman, 2003), which separately forecast the length of zero runs and the size of non-zero demands. Since the median length of a run of zeroes in our sample series is 1 week, with 90% of the zero runs are of four weeks' length or less, and since the candidate models as tested will accommodate runs of zero observations, we have not chosen to model the series as intermittent demands. It is possible, however, that the work here might contribute to the development of forecasting techniques for intermittent demand, and *vice versa*—see Section 5 for further discussion.

A number of researchers have examined issues arising when reporting forecasts for low count series. The assumption of normally-distributed observations in conventional forecasting practice makes forecast summaries particularly easy to produce. Since the mean, median and mode of the normal distribution all coincide, it makes sense to report a point forecast in a variety of contexts (with, for example, quadratic or absolute loss functions). And since the normal distribution is symmetric about its mean, the standard error provides a convenient indication of forecast uncertainty.

Forecasts for low count series, unfortunately, are by not so neatly characterized. As McCabe and Martin (2005) point out, the mean of a predictive distribution for low counts is rarely itself an integer, so that reporting it as a forecast is "incoherent," since no future observation will ever match a non-integer forecast. And though the predictive median usually (though not invariably) constitutes a coherent forecast, it need not comprise the most likely future observation. Similar problems attach to the reporting of forecast uncertainty; for example, a confidence interval derived from a skew distribution gives no indication as to where predicted observations are likely to be found in the interval. To ameliorate such concerns, McCabe and Martin (in common with Willemain, 2006) advocate using the full predictive distribution as a forecast. As we pointed out above, Bayesian forecasting using a state-space model produces such full distributions as a matter of course.

The use of predictive distributions in forecasting is actually an established practice in the field of operational meteorology (see Wilks, 1995), where they are referred to as *probabilistic forecasts*—a term we adopt here, along with the techniques for forecast appraisal described in Section 4.

3 Candidate Models

Four models are actually compared in this paper—three state space models, and one "baseline" model, intended to represent a forecasting method that requires minimal effort to implement. Structurally, all three state space models are of the simplest possible type: the *local level*. This means that the state vector in each case consists simply of a single variable subject to a random walk over time. Certainly it's possible that more elaborate models might be better fitted to some of the time series in the test sample, but our intent in the exercise was to mimic the context of most inventory planning for low-demand items; in most cases, it's uneconomic to devote more than the barest manual effort to any individual item, and so simple generic forecasting techniques tend to be applied to all of them. The state space models tested are all structurally analogous to the sort of simple exponential smoothing widely employed for inventory planning (Gardner, 1985), and differ only in their stochastic characteristics. A minor technical impediment to the comparison of more structurally sophisticated approaches is that one of our candidate models (*GP* below) currently exists in the literature only in local level form. Though Harvey (1989) does illustrate the use of exogenous covariates with the model, it's by no means guaranteed that suitable covariates would be available to a forecaster in the context we seek to emulate. The test models are introduced in turn in the remainder of this section.

Adjusted Gaussian Dynamic Linear Model (AG)

The first of our test models challenges the assertion in Section 2 that forecasting low count series requires specialized models. This model is no more than a conventional Gaussian local level model as described in (West and Harrison, 1997). (We should note that West and Harrison do not necessarily recommend such a model for low count series.) In this model, the time series Y_t to be forecast is normally distributed around a latent level series λ_t, with observation noise of variance ν^2. The latent series itself follows a random walk, perturbed in each period by evolution noise with variance ω^2:

$$Y_t \sim \mathrm{N}(\lambda_t, \nu^2)$$
$$\lambda_t \sim \mathrm{N}(\lambda_{t-1}, \omega^2)$$

In general, of course, the forecast distributions from such a model are incoherent for low count series, since they encompass non-integer and negative values which will never equal an observed value of the series. To address this problem, we adjust the model's predictions as follows: Assume that the forecast series may take distinct count values in the set $\{y_1, \ldots, y_K\}$ (for the demand series in the sample, this is the set of integers $\{0, \ldots, 9\}$). Let $p^*(Y_h)$ be the (incoherent) forecast distribution produced directly by the model at forecast horizon h. The probabilistic forecast actually reported consists of the vector $p_{1h}, \ldots, p_{Kh}$ where:

$$p_{kh} = c_h p^*(y_k - 0.5 < Y_h \leq y_k + 0.5), \quad \text{for } k = 1, \ldots, K$$

Here, c_h is a normalizing constant chosen so that $\sum_{k=1}^{K} p_{kh} = 1$. *Ad hoc* though this approach might appear, it parallels the established practice of using a linear-Gaussian forecasting model, trimming and rounding the predictive mean and/or quantiles it produces. Bayesian forecasting with the AG model may be achieved using well-known simulation algorithms, as discussed in (West and Harrison, 1997) and (Durbin and Koopman, 2001).[1] The assumption of normal distributions for both observation and evolution errors allows the Kalman filter to be applied directly to produce closed-form updates, greatly expediting the estimation process.

Poisson Dynamic Log-Linear Model (PL)

The second candidate model is structurally similar to the first. As in the first model, the latent level evolves according to a random walk with normal noise of variance ω^2. In this case, however, observations are distributed according to a Poisson distribution with mean equal to the exponential of the latent level:

$$Y_t \sim \mathrm{Poisson}(\exp \lambda_t)$$
$$\lambda_t \sim \mathrm{N}(\lambda_{t-1}, \omega^2)$$

The latent level process may be considered as a dynamic analog of the linear predictor of a log-linear model (Fahrmeir and Tutz, 1994). In contrast to those of the first model,

[1] Software to estimate all the models described in this paper (written in the statistical programming language R), together with the data used in the evaluation, are available from the author on request.

predictions produced from this model are coherent for low count series, and may be used without further alteration. The use of a Poisson distribution in this model precludes direct use of the Kalman filter, resulting in a more computationally expensive estimation process. For our purposes, however, we have found the straightforward simulation approach of Tanizaki and Geweke (1999) to be sufficiently speedy.

Gamma-Poisson Local Level Model (GP)

This model, presented in (Harvey and Fernandes, 1989) and (Harvey, 1989), is a state space model with closed form updating like the AG model that produces coherent count forecasts like the PL model. As in the PL model, observations in this model follow a Poisson distribution conditional on the underlying level. In this model, however, the level itself is used, not its exponential:

$$Y_t \sim \text{Poisson}(\lambda_t)$$

The latent level here follows a series of conditional gamma distributions, rather than the normal distributions of the first two models. This series begins with a distribution which we represent as a "posterior distribution" for a notional level value λ_0, with an empty conditioning set:

$$\lambda_0 \mid \emptyset \sim \text{Gamma}(a_0, b_0)$$

The conditional prior for level $\lambda_{t\geq 1}$ is then given inductively:

$$\begin{aligned} \lambda_{t-1} \mid Y_1, \ldots, Y_{t-1} &\sim \text{Gamma}(a_{t-1}, b_{t-1}) \quad \Rightarrow \\ \lambda_t \mid Y_1, \ldots, Y_{t-1} &\sim \text{Gamma}(wa_{t-1}, wb_{t-1}) \end{aligned}$$

Here, w is a model parameter such that $0 < w \leq 1$. Once the value Y_t has been observed, standard results for Bayesian inference imply that the posterior distribution for λ_t is also a Gamma distribution:

$$\lambda_t \mid Y_1, \ldots, Y_{t-1}, Y_t \sim \text{Gamma}(wa_t + Y_t, wb_t + 1)$$

Estimation of this model and concomitant forecasting may be carried out by embedding the updating formulas above in a simulation algorithm.

Climatological Baseline Model (Cm)

For our baseline model, we follow established practice in meteorology, where the baseline model for probabilistic forecasting is termed the *climatological model*, and simply forecasts events (at all horizons) with probabilities equal to their observed marginal frequencies. A formal Bayesian embodiment of such a model might comprise a multinomial distribution for the observed counts, governed by a Dirichlet distribution:

$$\begin{aligned} Y_t &= \textstyle\sum_{k=1}^{K} k \times Z_{tk} \\ Z_t &\sim \text{Multinomial}(1; \alpha_1, \ldots, \alpha_K) \\ (\alpha_1, \ldots, \alpha_K)' &\sim \text{Dirichlet}(\delta_1, \ldots, \delta_K) \end{aligned}$$

Here, Z_t is a vector of K elements consisting of $K-1$ zeros and 1 one. Forecasting with this model is trivial: If n_{kt} denotes the number of times we have observed $Y_{t'} = k$, for $t' \leq t,\ k = 1, \ldots, K$, the posterior distribution of Z_t is:

$$\begin{aligned} Z_t | Y_1, \ldots, Y_t &\sim \text{Multinomial}(1; \alpha_{1t}, \ldots, \alpha_{Kt}) \\ (\alpha_{1t}, \ldots, \alpha_{Kt})' &\sim \text{Dirichlet}(\delta_1 + n_{1t}, \ldots, \delta_K + n_{Kt}) \end{aligned}$$

The probabilistic forecast at t for any time horizon is given by the mean of the posterior parameter distribution, namely a vector $(p_{1t}, \ldots, p_{Kt})'$, where, for $k = 1, \ldots, K$:

$$p_{kt} = \frac{\delta_k + n_{kt}}{\sum_{j=1}^{K} (\delta_j + n_{jt})}$$

4 Test Results and Discussion

We tested our candidate models using the 100 series of parts demands introduced in Section 1. Each 78-week sample series was randomly divided into two series of at least 13 and at most 65 $(= 78 - 13)$ weeks. A candidate model was fitted each such sub-series and used to produce probabilistic forecasts with $1, \ldots, 13$ week horizons. The corresponding values from the original series were used to validate the forecasts. This procedure resulted in some $100 \times 2 \times 13 \times 4 = 10{,}400$ forecast-observation pairs—2,600 for each model, and 200 for each model at each forecast horizon. In order to compare the efficacies of the candidate models, we used two techniques from operational meteorology, which are outlined below.

Discrete Ranked Probability Scores

Described in (Murphy, 1969), the *discrete ranked probability score* (*DRPS*) summarizes the performance of a probabilistic forecasting method as a single number. The DRPS is closely related to another widely-used performance metric, the *Brier score* (defined in (Wilks, 1995), pp. 284 - 287, for example). Unlike the Brier score, however, the DRPS takes into account the ordinality of the variable being forecast (a forecast of 5 units, for instance, is considered worse than a forecast of 3 units when actual demand is 2 units), which is an appealing quality when dealing with predictions of demand.

To define the DRPS, consider a model producing probabilistic forecasts for a variable Y, the latter taking discrete values $y_1 < y_2 < \ldots < y_K$. Since Y is discrete, the forecast takes the form of a vector of probabilities $(p_1, p_2, \ldots, p_K)$, where p_k is the forecast probability that $Y = y_k$. We can calculate corresponding the cumulative forecast distribution $(P_1, P_2, \ldots, P_K)$, where $P_k = \sum_{j=1}^{k} p_j$ is the forecast probability that $Y \leq y_k$. With $\mathcal{I}(\cdot)$ denoting the indicator function whose value is 1 if its argument is true and 0 otherwise, define:

$$\Delta \stackrel{\text{def}}{=} \frac{1}{K} \sum_{k=1}^{K} (P_k - \mathcal{I}\{Y \leq y_k\})^2 \tag{1}$$

The DRPS of the model is simply the expectation of Δ, i.e. DRPS $\stackrel{\text{def}}{=} \mathrm{E}[\Delta]$. Informally, it is the average squared discrepancy between the cumulative forecast distribution and the (trivial) empirical distribution of the corresponding observation. Note that it is *negatively oriented*, in that a *lower* score reflects a more accurate forecast. If we use a model to make forecasts on a number of different occasions (not necessarily for the same time series), observing the corresponding actual values of Y, values of Δ may be calculated for each forecast and the average (the arithmetic mean, or the median) of these Δ's may used to approximate the DRPS of the model.

Table 1 displays the results of estimating the DRPS of each of candidate models over all forecasts horizons (i.e. all 2,600 test cases for each model), where both the median and arithmetic mean of the values of Δ are used as a proxy for its expectation. The rank of each model appears in parentheses next to their respective scores (recall that a *lower* score is better). In figure 3, the values of Δ for each model over all horizons are displayed in a box-plot, ranked in order of increasing median. A "notch" appears in each box at the median value; where notches do not overlap, the difference between the corresponding medians is statistically significant. (In fact, the Wilcoxon rank sum attests to a difference in medians significant at the 5% level between each of the models.) Finally, table 2 gives the median-derived DRPS value (and rank) for the models at time horizons $1, \ldots, 13$ (the corresponding mean values do not differ appreciably).

The implications of these results may be informally summarized as follows:

Model	DRPS (Rank)	
	Median	Mean
GP	0.861 (1)	1.32 (1)
PL	0.947 (2)	1.53 (3)
Cm	0.981 (3)	1.40 (2)
AG	1.293 (4)	1.69 (4)

Table 1. DRPS by model

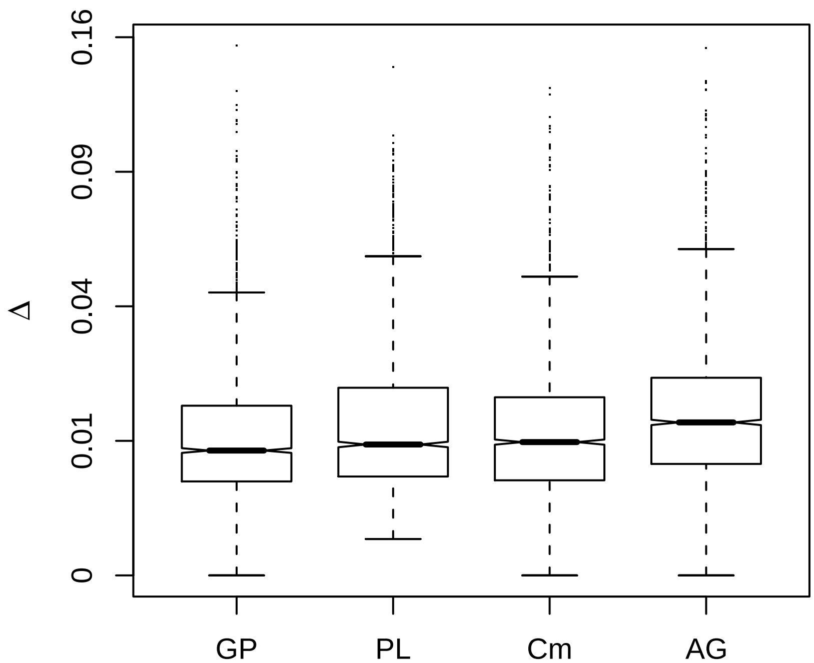

Figure 3. Box-and-whisker plot of Δ by model

- The GP model is the superior performer, both in the aggregate and at every forecast horizon.
- The performance of the AG is worst, again both in the aggregate and at every forecast horizon.
- Forecast produced by the PL model and the baseline Cm model are of roughly equal quality, with some indication that the PL model performs slightly better, though the difference is hardly overwhelming.

Horizon	Model (Rank)			
	Cm	AG	PL	GP
1	0.925 (3)	1.284 (4)	0.887 (2)	0.844 (1)
2	0.866 (2)	1.147 (4)	0.947 (3)	0.811 (1)
3	0.963 (3)	1.383 (4)	0.930 (2)	0.846 (1)
4	0.954 (3)	1.239 (4)	0.919 (2)	0.800 (1)
5	0.990 (3)	1.324 (4)	0.970 (2)	0.912 (1)
6	1.007 (3)	1.274 (4)	0.947 (2)	0.856 (1)
7	0.997 (2)	1.243 (4)	1.042 (3)	0.923 (1)
8	0.993 (3)	1.303 (4)	0.937 (2)	0.800 (1)
9	0.987 (3)	1.270 (4)	0.944 (2)	0.817 (1)
10	1.070 (3)	1.332 (4)	1.021 (2)	0.915 (1)
11	0.930 (2)	1.348 (4)	1.010 (3)	0.874 (1)
12	1.057 (3)	1.387 (4)	1.009 (2)	0.919 (1)
13	0.941 (2)	1.334 (4)	0.955 (3)	0.820 (1)

Table 2. Median DRPS by model and forecast horizon

Multicategory Reliability Diagrams

For a more detailed assessment of forecast performance, we turn to the *multicategory reliability diagram* or *MCRD* introduced in (Hamill, 1997). A variant of the widely-used *reliability diagram* described in (Wilks, 1995), the purpose of the MCRD is to illustrate the *calibration* of a forecasting method—that is, the degree to which events forecast with a certain probability actually occur with that probability. It displays, for a prespecified set of quantiles, the fraction of observations that fall at or below the corresponding quantile of the forecast distribution. Formally, we begin by converting the probabilistic forecast $p_1, \dots, p_K$ into a set of quantile values $z_1, \dots, z_M$, corresponding to quantile positions $q_1, \dots, q_M$, where for $q \in \{q_1, \dots, q_M\}$:

$$z_q = \min\{k \mid \left(\sum_{j=1}^{k} p_j\right) \geq q\}$$

With a perfectly calibrated forecasting method for a continuous variable, we would expect on average 25% of observations should be below the 25th percentile of the forecast distribution, 75% below the 75th percentile, and so on. Of course, with a discrete predictand such as low count demand, it's possible that more than one quantile of the forecast distribution will have the same value; to cope with this, Hamill defines the "interpolated probability" that the predictand Y is less than the quantile value z_q as follows:

$$\widetilde{P}(Y < z_q) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } \; Y > z_q \\ (q - q_{\min})/(q_{\max} - q_{\min}) & \text{if } \; Y = z_q \\ 1 & \text{if } \; Y < z_q \end{cases}$$

Here, $q_{\max}$ and $q_{\min}$ are respectively the maximum and minimum of the set $\{q' \,|z_{q'} = Y\}$, for $q' \in \{q_1, \ldots, q_M\}$. The *calibration*, C_q, at quantile position q is the expected interpolated probability that Y is less than z_q:

$$C_q \stackrel{\text{def}}{=} \mathrm{E}\left[\widetilde{P}(Y < z_q)\right] \tag{2}$$

As with the DRPS, we can use the mean or median of the values across all the test cases to estimate the expectation in the above; since the distribution of C_q turns out to be reasonably symmetric, we've used the mean in the following, though the differences that result from using the median are insignificant. Figure 4 displays the MCRD of each of the candidate models (for comparison, the remaining models are shown on each panel in gray). We've calculated calibration values at percentile positions $5, 10, ..., 95$. In a perfectly calibrated forecast, $C_q = q$, so the deviation of a plot from the diagonal (also displayed on the graphs) is an indication of forecast errors made by the model.

Inspecting the figure, it appears that all of the candidate models are *over*-forecasting, in that the forecast distribution quantiles are greater than the corresponding quantiles of the observation distribution (more than 50% of the observations, for example, are less than the 50^{th} percentile of each model's forecast distributions). The most egregious over-forecaster is the AG model, as might be expected from its poor DRPS scores. The GP model, though it also over-forecasts, departs less from the diagonal (i.e. perfect calibration) than the other models at every quantile. The PL model looks to be better calibrated than the Cm model at the lower quantiles (generally, lower count values), but over-forecasts at higher quantiles, so it would seem that the PL model is producing forecast distributions with unwarrantedly heavy right tails.

5 Conclusion

From the preliminary investigations described in this paper, it would appear that in this application at least, the gamma-Poisson local level model is the superior forecaster, significantly improving upon the performance of the baseline model, and of the dynamic log-linear model, too. By contrast, ad-hoc adjustment of a linear-normal state space model produces forecasts that are inferior not only to the specialized state space

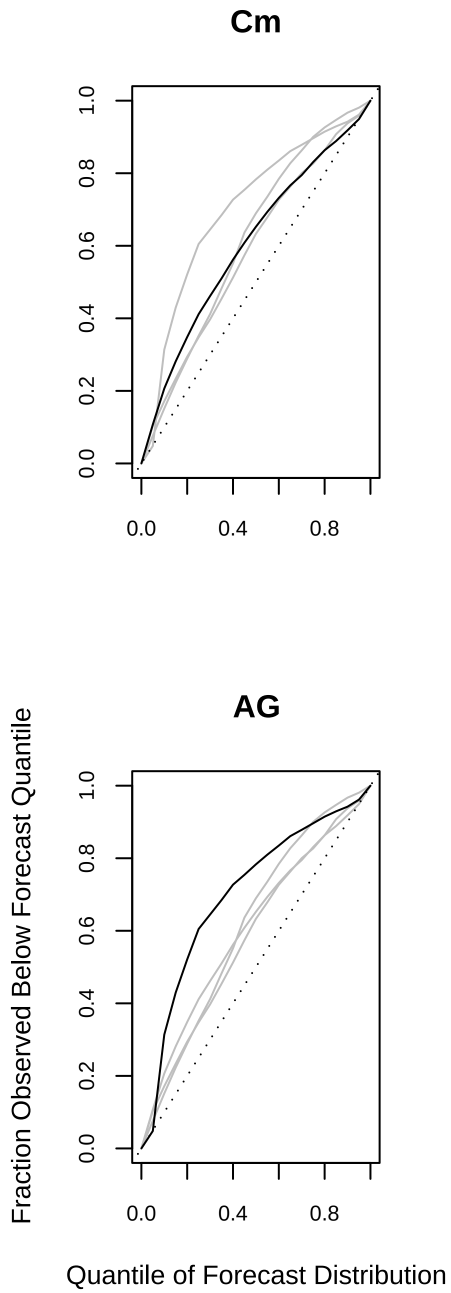

Figure 4. Multicategory Reliability Diagrams by Model

models, but to the baseline model as well. The latter results should sound a note of caution regarding the routine use of conventional exponential smoothing methods to forecast inventory demands of the sort examined here. The performance of the dynamic log-linear model (which barely improves upon the baseline) is somewhat disappointing; the prospect of employing a direct analog of generalized linear models in this application is an appealing one, and significant effort has been expended on their efficient

Bayesian estimation (see de Jong and Shephard, 1995 or Durbin and Koopman, 2001, for example). We are investigating whether the performance of the PL model might be improved, by using a non-canonical link (rather than the exponential used here) or by imposing an informative prior on the evolution variance, to reduce the right tails of the model's predictive distributions.

As we noted in the previous section, all of the candidate models tend to produce forecasts in excess of the actual demands. Two explanations for this tendency—and consequent remedies—suggest themselves: (1) Though Figure 2 shows that all demands are bounded to 9 units or less, all of the state space models considered here have essentially unbounded level components. We are currently working to establish whether models with finite discrete state spaces (McDonald and Zucchini, 1997) produce less inflated forecasts in situations where a definite bound can be placed on all future demands. (2) Case-by-case examination of the forecasts made during the comparison exercise indicates that all of the models fail to anticipate phases of a part's lifecycle—particularly the diminution in demand that accompanies phase-out. Unfortunately, given the highly decentralized nature of Sun's operations, basic facts such as part end-of-life dates are rarely collected together in a systematic fashion, making it far from trivial to predict the commencement of a part's phase-out. We are, however, exploring more sophisticated model structures built on the work described in (Yelland, 2004), to see if it is possible statistically either to predict or rapidly detect transitions in part lifecycles.

Finally, though we observed that long periods without demand occur relatively rarely in the sample series, the distinction between "low count" and "intermittent" demands is not a sharp one, and it's quite reasonable to consider inclusion of forecast model components specifically to address demand intermittance. Since the latter normally center on estimating the length of breaks in demand, and since such breaks are themselves likely to be low counts in many instances, it's also possible that the results of our investigations here might be useful in the development of forecasting techniques for intermittent demand, too.

References

Al-Osh, M. A., Alzaid, A. A., June 1990. An integer-valued pth order autoregressive structure (INAR(p)) process. Journal of Applied Probability 27 (2), 314–324.

Boylan, J., June 2005. Intermittent and lumpy demand: A forecasting challenge. Foresight: The International Journal of Applied Forecasting 1, 36–42.

Cameron, A. C., Trivedi, P. K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge.

Clement, J., Coldrick, A., Sari, J., 1995. Manufacturing Data Structures: Building Foundations for Excellence with Bills of Materials and Process Information. Wiley, New York.

Cox, D. R., 1981. Statistical analysis of time series: Some recent developments. Scandinavian Journal of Statistics 8, 93115.

Croston, J. D., 1972. Forecasting and stock control for intermittent demand. Operational Research Quarterly 23, 289–303.

de Jong, P., Shephard, N., June 1995. The simulation smoother for time series models. Biometrika 82 (2), 339–350.

Durbin, J., Koopman, S. J., 2001. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

Fahrmeir, L., Tutz, G., 1994. Multivariate Statistical Modeling Based on Generalized Linear Models. Springer-Verlag, New York.

Gardner, E. S., 1985. Exponential smoothing: The state of the art. Journal of Forecasting 4 (4), 1–28.

Hamill, T. M., 1997. Reliability diagrams for multicategory probabilistic forecasts. Weather and Forecasting 12, 736–741.

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

Harvey, A. C., Fernandes, C., 1989. Time series models for count or qualitative observations. Journal of Business and Economic Statistics 7 (4), 407–417.

McCabe, B., Martin, G., 2005. Bayesian predictions of low count time series. International Journal of Forecasting 21 (2), 315–330.

McDonald, I. L., Zucchini, W., 1997. Hidden Markov and Other Models for Discrete-Valued Time Series. Chapman and Hall, London.

McKenzie, E., 2003. Discrete variate time series. In: Shanbhag, D., Rao, C. (Eds.), Handbook of Statistics. Vol. 21. Elsevier, Amsterdam, pp. 573–606.

Murphy, A. H., 1969. On the ranked probability score. Journal of Applied Meteorology 8, 988–989.

Shenstone, L., Hyndman, R. J., February 2003. Stochastic models underlying Croston's method for intermittent demand forecasting. Monash Econometrics and Business Statistics Working Papers 1/03, Monash University, Department of Econometrics and Business Statistics.

Tanizaki, H., Geweke, J., 1999. On Markov chain Monte Carlo methods for nonlinear and non-Gaussian state-space models. Communications in Statistics, Simulation and Computation 28 (4), 867–894.

West, M., Harrison, P. J., 1997. Bayesian Forecasting and Dynamic Models, 2nd Edition. Springer-Verlag, New York.

Wilks, D. S., 1995. Statistical Methods in the Atmospheric Sciences: An Introduction. Academic Press, New York.

Willemain, T., June 2006. Forecast-accuracy metrics for intermittent demands: Look at the entire distribution. Foresight: The International Journal of Applied Forecasting 4, 36–38.

Winkelmann, R., 1997. Econometric Analysis of Count Data, 3rd Edition. Springer-Verlag, New York.

Yelland, P. M., 2004. A model of the product lifecycle for sales forecasting. Tech. Rep. 127, Sun Microsystems Laboratories.
URL http://research.sun.com/techrep/2004/smli_tr-2004-127.pdf