Using Customers' Reported Forecasts to Predict Future Sales

Nihat Altintas †, Alan Montgomery ‡, Michael Trick ‡

Tepper School of Business,
Carnegie Mellon University, Pittsburgh, PA 15213.
†nihat@andrew.cmu.edu
‡{alan.montgomery,trick}@cmu.edu

May 12, 2006

Abstract

We consider a supplier who must predict future orders using forecasts provided by their customers. Our goal is to improve the supplier's operations through a better understanding of the customers's forecast behavior. Unfortunately, customer forecasts cannot be used directly. They may be biased since the customer may want to mislead the supplier into believing that orders may be larger than expected to secure favorable terms or simply because the customer is a poor forecaster. There are several unusual elements in our problem. Analysts typically observe the actual process which may be biased due to asymmetric loss function. Also units are discrete not continuous. We believe the forecasting process can be modeled in a multi stage process. The customer first computes the distribution of demand at a future date which we assume to follow an ARMA(1,1) model. The customer then submits an integral forecast(multiples of lot sizes), which minimizes his expected loss due to forecast errors. We show that the customer may bias his forecasts under an asymmetric loss functions. We provide an estimator which can be used by the supplier to estimate the forecast generation model of the customer by looking at his forecasts.

1 Introduction

In the last two decades, business environments are undergoing rapid changes due to expansion of computer resources. Companies can now share huge amount of information across their supply chains. Empowered with the information, companies have a large spectrum of choices to create complex products and processes. Due to increasing complexity, we observe more collaborative effort between the different parties in the supply chain. To achieve higher returns, suppliers and customers invest in technologies that provide real-time access to demand, inventory, price, sourcing, and production data. Sharing information is key to increase the profits under demand uncertainty. A critical factor that determines the quality of information transmission is its reliability. Inaccurate information can lead to severe costs for the parties in the supply chain. Most of the existing methods are not appropriate because of the potential bias in the forecasts. However, the suppliers still collect forecasts from the customers and do not only use their own forecasts. The biased forecasts still contain information which might be useful for the supplier. It is critical for the supplier to detect the bias and predict the future order of a customer.

We provide an empirical study about forecast sharing in the supply chains. We analyze the forecasts of final orders that are received by an automotive supplier who produces multiple parts for auto manufacturers. Forecast information helps the supplier to predict the final orders to do production planning in advance. As part of the collaborative effort, customers provide forecast updates to the supplier at different order dates.

Example 1 *Forecasts updates for 1-Aug, 8-Aug and 15-Aug*

1-Aug		**8-Aug**		**15-Aug**	
Due Date	**Quantity**	**Due Date**	**Quantity**	**Due Date**	**Quantity**
8-Aug	*800*				
15-Aug	*800*	*15-Aug*	*160*		
22-Aug	*960*	*22-Aug*	*480*	*22-Aug*	*320*
29-Aug	*960*	*29-Aug*	*640*	*29-Aug*	*640*
5-Sep	*960*	*5-Sep*	*800*	*5-Sep*	*800*
		12-Sep	*960*	*12-Sep*	*800*
				19-Sep	*960*

Example 1 represents a typical forecast for a customer. The customer provides different updates in August 1^{st}, August 8^{th} and August 15^{th}. For example in forecast date August 1^{st}, the customer provides forecasts for the next 5 weeks with quantities 800, 800, 800, 960 and

960. In each forecast date, the customer updates the forecasts from the previous forecast date and can place new orders. Therefore, the forecasts can be considered as a flow of orders which evolve over time. As can be noticed from the forecast values, due to production and transportation constraints the customer provides forecasts which are multiples of some lot sizes. Therefore, we can divide all the forecast by a common divisor (160) to obtain the number of batches in each order. In this analysis, we propose a framework for modeling the forecast generation process at the customer.

Forecasts provide information about the future orders of the customer. However, the forecast information can be quite noisy and can be misleading for the supplier. It is critical for the supplier to process the forecasts to estimate future orders. Altintas and Trick (2004) study a similar data set in a non-parametric framework and provide empirical support for downstream players consistently over- or under-estimating their forecasts through time. It is important for a supplier to recognize a significant pattern in the forecasts by looking at order history of a customer. Our objective here is to provide a framework to extract information from the forecast data and adjust the forecasts to provide a better estimate of future orders. For example, if a customer constantly overestimates his orders, then the supplier can detect this behavior and can remove the bias from the forecast.

Why do customers provide poor forecast performance?

There are a couple factors that can lead to forecast errors.

1. **Uncertainty in the usage:** There is always uncertainty in the customer's system due to demand variance, lead times, machine failures and etc. Therefore, the error in the forecast can be a result of these factors. As the due date approaches, the customer has more information about the demand and uncertainty decreases.

2. **Bias in the forecast:** The customer can have different costs for overestimation and underestimation. When the customer overestimates, the supplier can penalize the customer or the customer can lose the goodwill of the supplier. In the case of underestimation, the customer cannot satisfy the demand which can lead to delays in production and potentially lost sales. The customer must consider the trade off between overestimation and underestimation and submit a forecast which minimizes the expected cost. The unit overestimation cost is not necessarily equal to underestimation cost, therefore the customer may add a bias to his forecast to minimize his cost.

Therefore a realistic model should consider the uncertainty in production and bias in the forecast.

1.1 Literature Review

Forecasting has been addressed in many different problem settings. There is huge body of literature in forecasting new observations. In our analysis, the suppliers collect self-reported forecasts which might be biased from the customers. We use a bayesian approach to estimate the model parameters in our analysis. Geweke and Whiteman (2006) provide an extensive review of literature in bayesian forecasting models. Another stream of research which deals with self-reported forecasts is consensus forecasting models. Batchelor and Dua (1995)) show that combinations of different forecasts even from a small number of sources is helpful in predicting future values. In our case, we have regular updates from the customers which are combined to predict the customer demand. Therefore, the final prediction incorporates different updates of the customer.

In our analysis, we model the demand function as a time-series process. The customer submits his final forecast after considering the cost of overestimation and underestimation. Time-series assumption has been also used by some other researchers to provide theoretical results about forecasting. Graves (1999) assumes a non-stationary demand with ARIMA process and shows that inventory decisions behaves much differently compared to a stationary process. Aviv (2003) proposes a unified time-series framework for forecasting and inventory control. He assumes that different parties observe different subsets of information, and adopt their forecasting and demand process accordingly. Aviv (2001, 2002) also study models with time-series demand. Our results provide an empirical support for forecasting models which assume strategic behavior of parties with private information in a time-series framework. Chen et al. (2000a, 2000b) assumes that downstream player uses moving average forecasts to place orders to a supplier. They measure the amplification in the variance of the orders which is known as bullwhip effect.

There are other mathematical models in literature for the evolution of demand. Graves et al. (1986a, 1986b, 1998) and Heath and Jackson (1994) develop the Martingale Model of Forecast Evolution (MMFE) to model the evolution of forecasts. In MMFE, a forecaster creates forecasts for the planning horizon and update them in regular intervals. The error of the forecast updates are assumed to follow the Martingale Property: independent, identically distributed, multivariate normal random variables with mean 0. MMFE approach has been studied by a number of researchers under different problem settings. (see, e.g., Gullu (1996) , Graves et al. (1998), and Toktay and Wein (2001)). Another approach is considering that some demand parameters are unknown in advance and using Bayesian updates to incorporate new information as it becomes available (Scarf (1959) , Azoury (1985) and Lariviere and Porteus (1999))

Our analysis has several aspects that have not been addressed in the literature using self-reported forecasts. First, we assume a loss function for the forecast errors and explain the dynamics behind the loss function by using a time-series demand. Our analysis is the first to put the ARIMA form into a newsvendor framework with a cost of overestimation and underestimation. Second, we assume forecast values which are discrete values since they are multiples of lot sizes. Third, we provide a statistical procedure to estimate the model parameters for this complicated problem. A limitation is that we will not get into a symmetric game, where customer adjusts to the suppliers and vice versa.

The empirical research about forecasting is very limited in the supply chain literature. Terwiesch et al. (2003) considers the problem from a buyer's perspective. They demonstrate that poor forecast performance, in terms of forecast inflation and volatility, damages the buyer's reputation and leads to a lower service. In our analysis, we looked at the problem from the supplier's perspective and provide analysis to understand the forecast behavior of the customers.

2 Problem Environment

In our analysis, we model the orders that are placed to an automotive parts supplier by auto customers. Customers place some preliminary orders (forecasts) starting from six month before the order date and adjust the forecasts before the due date. The parts are engine systems and generate multi-billion dollar revenue for the supplier. The customer has a better ability to predict the demand due to its proximity to the final demand. The supplier can only observe the forecasts submitted by the customer. In Figure 1, we can see that the manufacturers provide different updates at 1-Aug, 8-Aug and 15-Aug for the next 5 weeks.

2.1 Demand Model

In our analysis, we observe that the forecasts and forecast errors are autocorrelated through time. between the periods. A forecast error leads to excess inventory or backordering and carried to the next period. Another important factor is the autocorrelation between the demand values of consecutive periods. For example, high demand periods can be followed with low demand. Therefore, the demand model should be fairly adaptive in order to incorporate the available information in each period. ARMA model provide a flexible model to describe different demand processes. (Box and Jenkins 1970) Let X_t as the demand of

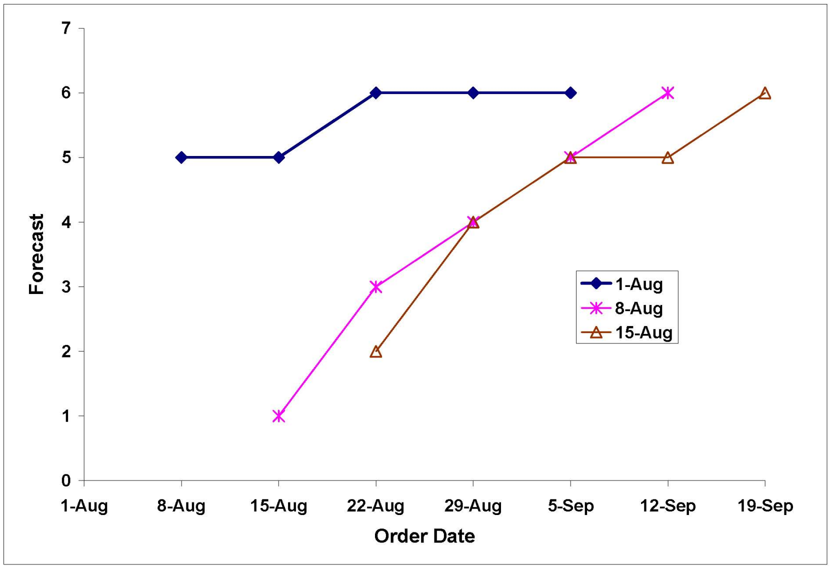


Figure 1: The forecast updates at different order dates 1-Aug, 8-Aug and 15-Aug.

the customer at time t. We model X_t as

$$X_t - \mu = \phi(X_{t-1} - \mu) + \epsilon_t - \theta\epsilon_{t-1} \tag{1}$$

$$X_t = (1-\phi)\mu + \phi X_{t-1} + \epsilon_t - \theta\epsilon_{t-1} \tag{2}$$

which is $ARMA(1,1)$ with $E(X_t) = \mu$. Having both autoregressive and moving average components, the $ARMA(1,1)$ model can capture the significant behaviors in the data. Our framework can be extended over higher number of lags. We assume that the error ϵ_t is normally distributed with $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \sigma^2$.

2.2 Forecast Generation Model

We assume that forecasts are generated in a three stage process:

i. At time t, the customer considers the next K periods based on his available information set Ω_t. He computes the distribution $F(X_{t+1}, ..., X_{t+K}|\Omega_t)$ for the next K periods.

ii. The customer derives his optimal forecasts $\bar{X}_t = (\bar{X}_t(1), ..., \bar{X}_t(K))$ using an asymmetric loss function which represents the strategic behavior of the customer.

iii. At time t, the customer submits a forecast vector to the supplier, $\hat{X}_t = (\hat{X}_t(1), ..., \hat{X}_t(K))$ where $\hat{X}_t(k) = \bar{X}_t(k) + \gamma_t(k)$ for $k = 1, ..., K$. Errors $\gamma_t(k)$ might perturb the process to generate non-ARIMA forecasts.

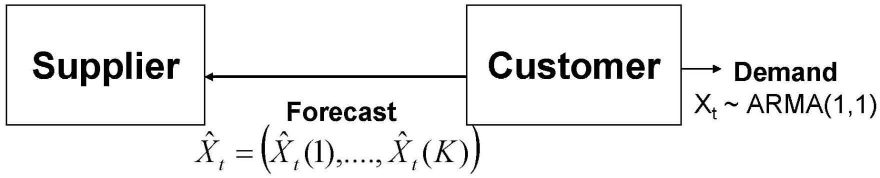


Figure 2: Manufacture first observes demand and submit a forecast to the supplier

The customer observes demand which takes continuous values. However, from Example 1, we observe that the customer places integral forecasts which are multiples of lot sizes. Therefore, we assume that the forecasts can only take integer values. Next we explain each step of the forecast generation model in detail.

2.2.1 Derivation of $F(X_{t+1},...,X_{t+K}|\Omega_t)$ at time t

By using $ARMA(1,1)$, we also model the uncertainty in demand for future periods. At time t, the customer can derive the k-step ahead demand X_{t+k} as

$$X_{t+1} = (1-\phi)\mu + \phi X_t + \epsilon_{t+1} - \theta\epsilon_t \tag{3}$$

for $k = 1$ and

$$X_{t+k} = (1-\phi^k)\mu + \phi^{k-1}\left(\phi X_t - \theta\epsilon_t\right) + \sum_{n=1}^{k-1}\epsilon_{t+n}\phi^{k-n-1}(\phi-\theta) + \epsilon_{t+k} \tag{4}$$

for $k = 2,...,K$. Since the unobserved values $\epsilon_{t+1},...,\epsilon_{t+k}$ are Normal$(0,\sigma^2)$, we can also show that X_{t+k} is normally distributed with

$$\begin{aligned} E(X_{t+k}|\Omega_t) &= (1-\phi^k)\mu + \phi^{k-1}\lambda_t & k = 1,...,K \\ Var(X_{t+k}|\Omega_t) &= \sigma^2\upsilon_k^2 & k = 1,...,K \end{aligned}$$

where $\lambda_t = (\phi X_t - \theta\epsilon_t)$, $\upsilon_1 = 1$ and $\upsilon_k = \sqrt{(1+(\phi-\theta)^2\left(\frac{1+\phi^{2(k-1)}}{1-\phi^2}\right)}$).

2.2.2 Loss function

The customer may have private information which is not available to the supplier. The customer provides forecasts to minimizes his own cost. We model this strategic behavior as an asymmetric loss function with different values for costs of overestimation and underestimation. Therefore, the customer does not necessarily submit the expected demand as his forecast and can add a bias to his forecast. We provide an empirical framework to test the significance of this hypothesis by looking at the forecast data.

The customer can evaluate the cost of overestimation and underestimation to adjust his final forecast. A common assumption in supply chain management literature is having a linear cost function with c_o for each unit of overestimation and c_u for each unit of underestimation (Cachon 2004). These costs can be considered as the imputed cost of forecast errors. Cost of overestimation can include the penalty cost and goodwill of the supplier. The cost of underestimation can be composed of the expected backordering cost or lost sales. By following the same approach, we can write the loss function as

$$U(X_t(k), X_{t+k}) = \begin{cases} c_o|X_t(k) - X_{t+k}| & \text{if } X_t(k) \geq X_{t+k} \\ c_u|X_{t+k} - X_t(k)| & \text{otherwise} \end{cases}$$

where $c_o, c_u > 0$.

The customer can derive the expected loss (or risk) for any forecast $X_t(k)$ as follows

$$E(U(X_t(k), X_{t+k})|\Omega_t) = c_o \int_{-\infty}^{X_t(k)} (X_t(k) - X_{t+k})\, dF(X_{t+k}) + c_u \int_{X_t(k)}^{X_t=\infty} (X_{t+k} - X_t(k))\, dF(X_{t+k})$$

2.2.3 Submission of the Forecast

Therefore, at period t, the customer solves a minimization problem to minimize his expected loss function $E(U(X_t(k), X_{t+k}))$ and finds an optimal forecast

$$\bar{X}_t(k) = \text{argmin } E(U(X_t(k), X_{t+k})). \tag{5}$$

If we assume continuous valued forecasts, we can take the first derivative of the expected loss with respect to $X_t(k)$ to find the optimal continuous valued forecast $\zeta_t(k)$,

$$\begin{aligned} \frac{\partial E(U(X_t(k), X_{t+k})|\Omega_t)}{\partial X_t(k)} &= c_o \int_{-\infty}^{X_t(k)} X_{t+k} dF(X_{t+k}) - c_u \int_{X_t(k)}^{X_t=\infty} X_{t+k} dF(X_{t+k}) \\ &= c_o F(X_t(k)|\Omega_t) - c_u \left(1 - F(X_t(k)|\Omega_t)\right) \\ &= F(X_t(k)|\Omega_t)(c_o + c_u) - c_u \end{aligned}$$

Since $\frac{\partial^2 E(U(X_t(k), X_{t+k})|\Omega_t)}{\partial X_t(k)^2} = f(X_t(k)|\Omega_t) > 0$, from the first order conditions $\zeta_t(k)$ satisfies,

$$\begin{aligned} F(\zeta_t(k)|\Omega_t)(c_o + c_u) - c_u &= 0 \\ F(\zeta_t(k)|\Omega_t) &= \frac{c_u}{c_o + c_u} \end{aligned}$$

Therefore, $\zeta_t(k)$ for the customer is

$$\zeta_t(k) = F^{-1}\left(\frac{c_u}{c_o + c_u}|\Omega_t\right). \tag{6}$$

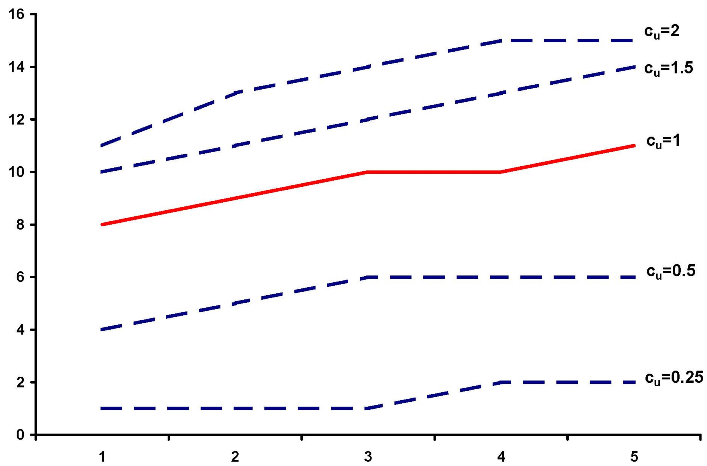


Figure 3: The optimal k-step ahead forecasts at time t for $\mu = 12$, $\sigma = 8$, $\phi = 0.7$, $\theta = 0.1$, $X_t = 6$, $\epsilon_t = 3$. The solid line represents an unbiased forecast with $c_u = c_o = 1$.

When the loss function is symmetric ($c_u = c_o$), the customer finds the median $F^{-1}(0.5|\Omega_t)$ as the optimal forecast. If the distribution is symmetric, then median is equal to mean and the customer submits $E(X_{t+k}|\Omega_t)$ as the optimal forecast. As can be seen in Figure 3 the customer adds a bias to his forecast when the the loss function is asymmetric. When $c_u > c_o$, the customer tends to overestimate his orders and adds a positive bias to his forecast. When $c_u < c_o$, it is less costly to underestimate for the customer. In this case the customer has a negative bias in his forecast. Therefore, asymmetric loss functions can explain the dynamics behind the bias in the forecasts.

When we assume the the ARMA(1,1) model, the k-period ahead optimal continuous valued forecast can be derived as :

$$\zeta_t(k) \quad = \quad (1-\phi^k)\mu + \phi^{k-1}\lambda_t + \tau\sigma\upsilon_k \quad k = 1, ..., K$$

where τ be the z-value of $\frac{c_u}{c_o+c_u}$ for standard normal distribution.

In our analysis, we assume that forecasts can only take integer values. Since $E(U(X_t(k), X_t(k)))$ is convex, the optimal integer valued forecast $\bar{X}_t(k)$ is

$$\bar{X}_t(k) = \begin{cases} \lfloor \zeta_t(k) \rfloor & \text{if } E(U(\lfloor \zeta_t(k) \rfloor, X_t(k))) < E(U(\lceil \zeta_t(k) \rceil, X_t(k))) \\ \lceil \zeta_t(k) \rceil & \text{otherwise} \end{cases}$$

where $\lfloor \cdot \rfloor$ is the floor function and $\lceil \cdot \rceil$ is the ceiling function.

We assume that there might be some errors $\gamma_t(k)$ added to the optimal integer valued forecasts to generate non-ARIMA forecasts. The final forecast is

$$\hat{X}_t(k) = \bar{X}_t(k) + \gamma_t(k)$$

where $\gamma_t(k)$'s are random integral errors. We assume that the forecasts in the errors are $\gamma_t(k)$ generated by a mixture of two discrete normal distributions

$$\gamma_t(k) \sim \begin{cases} N(0,\omega_1) & \text{with probability p} \\ N(0,\omega_2) & \text{with probability 1-p} \end{cases}$$

where p is the mixing probability.

The matrix form of the problem is available in Appendix A.2.

3 Supplier's Problem

At $t = T$, the supplier observes the forecast matrix $\hat{X}$ which consists of the forecast vectors $\left\{\hat{X}_1, ..., \hat{X}_T\right\}$ that are submitted at $t = 1, .., T$. By following the forecast generation model, at time t the customer submits a forecast based on the demand parameters (μ, ϕ, θ and σ), cost parameters (c_u and c_o) and recent information (X_t and ϵ_t). For the sake of simplicity we will denote the parameters as $\Lambda = (\mu, \phi, \theta, \sigma, c_u, c_o, X_0, \epsilon_0)$. When Λ and previous demand values X is available, it is straightforward to determine the forecast. However supplier cannot observe Λ and X. The supplier can only make inference about the demand and cost information of the customer by only looking at the forecast. Therefore, supplier's problem can be formulated as finding the distribution $P(\Lambda|\hat{X})$. The problem has the following complexities:

1. The final demand at the customer is assumed to be continuous. However, the customer provides integer forecasts which are multiples of production lot sizes.

2. The customer does not necessarily have a symmetric loss function. The customer's submits biased forecasts when $c_u \neq c_o$.

3. The supplier does not observe demand $X = (X_1, ..., X_T)$, nor the demand and cost parameters(Λ).

4. The forecasts can come from non-ARIMA process due to errors $\gamma_t(k)$. The supplier does not know $\Upsilon = (p, \omega_1, \omega_2)$ for $\gamma_t(k)$.

The supplier seems to have very limited information once we consider the complexity of the analysis. However, we provide a Bayesian model which can be used to make inference about the model parameters. Forecast vectors provide substantial amount information about the forecast generation model of a customer. At time t, the supplier observes forecasts for each of the following K periods. Therefore, the supplier has access to TK data points for T periods.

By using a Bayesian framework we first define priors for the model parameters:

$$\begin{vmatrix} \mu \\ \phi \\ \theta \\ \log(\sigma) \\ \log(c_u) \end{vmatrix} \sim N(\hat{\Lambda}, V)$$

$$\epsilon_0 \sim N(0,\sigma) \quad X_0 \sim N(\mu, \sigma\sqrt{\tfrac{1+\theta^2-2\theta\phi}{1-\phi^2}}) \quad p \sim Beta(a,b) \quad log(\omega_i) \sim N(0, v_i) \text{ for i=1,2}$$

Since the ratio of $\frac{c_u}{c_u+c_o}$ is important, we assume that $c_o = 1$.

Although we define priors for the parameters, we cannot still derive the distribution of $P(\Lambda|\hat{X})$. Therefore, we use gibbs sampling in order to make draws from Λ by defining X as a latent parameter. We iteratively sample

1. $\Lambda|X, \bar{X}$
2. $X, \bar{X}|\Lambda, \hat{X}, \Upsilon$
3. $\Upsilon|\bar{X}, \hat{X}$.

By using Bayes Formula, we can write the first step where we draw $\Lambda|X, \bar{X}$ as

$$P(\Lambda|X,\bar{X}) = \frac{P(\Lambda, X, \bar{X})}{P(X,\bar{X})} = \frac{P(X,\bar{X}|\Lambda)P(\Lambda)}{P(X,\bar{X})} \tag{7}$$

Since $P(X, \bar{X})$ is constant we have

$$P(\Lambda|X,\bar{X}) \quad \propto \quad L(\Lambda; X, \bar{X}) \cdot P(\Lambda) \tag{8}$$

where $L(\Lambda; X, \hat{X})$ is the likelihood of Λ for any given values of X and $\hat{X}$. $L(\Lambda; X, \hat{X})$ is zero if X and λ does not generate $\bar{X}$ as the optimal forecast. Therefore, the likelihood can be written as

$$L(\Lambda; X, \bar{X}) = \begin{cases} L(\Lambda; X) & \text{if } \Lambda \text{ and } X \text{ generates } \bar{X} \\ 0 & \text{otherwise} \end{cases}$$

where

$$L(\Lambda; X) = \left(\frac{1}{2\Pi\sigma^2}\right) exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\epsilon_t^2\right\} \tag{9}$$

4 Estimation

The supplier wants to derive the distribution of parameters given the forecast matrix, $P(\Lambda|\hat{X})$. Our sampler is a modified slice sampler with rejection. The general slice sampling algorithm (Neal 2003) is constructed using the principle that one can sample from a distribution by sampling uniformly from the region under the plot of its density function.

4.1 Mechanism of the Slice Sampler

Assume that we want to sample from a distribution for a variable $x \in R^n$, which has density function $f(x)$. We can introduce an auxiliary variable real variable, y and define joint distribution function $p(x, y)$ of x and y which is uniformly distributed over the region $S = \{(x, y) : 0 < y \leq f(x)\}$. S is the area under $f(x)$. Let $Z = \int f(x)$, then we have

$$p(x,y) = \begin{cases} 1/Z, & \text{if } 0 < y < f(x) \\ 0, & \text{otherwise} \end{cases}$$

The marginal density for x is

$$p(x) = \int_0^{f(x)} 1/Z dy = f(x)/Z \tag{10}$$

We can sample jointly for (x, y) and keep x values to replicate $f(x)$.

4.2 Our Sampler for Estimation

We first find an initial set of Λ, X, $\bar{X}$ and υ which can generate the forecast matrix $\hat{X}$. We then iteratively sample

1. $\Lambda|X, \bar{X}$
2. $X, \bar{X}|\Lambda, \hat{X}, \Upsilon$
3. $\Upsilon|\bar{X}, \hat{X}$.

Our sampler works as follows:

Iteration 0 (Initialization):

1. Draw $\mu^{(0)}$,$\phi^{(0)}$, $\theta^{(0)}$, $\sigma^{(0)}$, $c_u^{(0)}$, $\epsilon_0^{(0)}$ and $X_0^{(0)}$ from the prior distributions.

$$\left|\begin{array}{c} \mu^{(0)} \\ \phi^{(0)} \\ \theta^{(0)} \\ \log(\sigma^{(0)}) \\ \log(c_u^{(0)}) \end{array}\right| \sim N(\hat{\Lambda}, V)$$

$\epsilon_0^{(0)} \sim N(0,\sigma) \quad X_0^{(0)} \sim N(\mu, \sigma\sqrt{\frac{1+\theta^2-2\theta\phi}{1-\phi^2}})$
where $-1 \leq \phi^{(0)} \leq 1$ and $-1 \leq \theta^{(0)} \leq 1$.

2. Draw $\epsilon^{(0)} = \left\{\epsilon_1^{(0)}, ..., \epsilon_T^{(0)}\right\}$ from $N(0, I\sigma^2)$. Compute $X^{(0)}$ and optimal $\bar{X}^{(0)}$.

$$X_t^{(0)} = (1-\phi^{(0)})\mu^{(0)} + \phi^{(0)} X_{t-1}^{(0)} + \epsilon_t^{(0)} - \theta\epsilon_{t-1}^{(0)}$$

and

$$\bar{X}_t(k)^{(0)} = \text{argmin } E(U(X_t(k), X_{t+k})). \tag{11}$$

for $t = 1, ..., T$ and $k = 1, ..., K$.

3. Draw $p^{(0)}$, $\omega_1^{(0)}$ and $\omega_2^{(0)}$ from the priors

$p^{(0)} \sim Beta(a,b) \quad log(\omega_1^{(0)}) \sim N(0, v_1) \quad log(\omega_2^{(0)}) \sim N(0, v_2)$

Repeat the following for a specified number (I) of iterations. (for i=1,...,I)

Iteration i:

1. In this step, we generate $\Lambda^i | X^{(i-1)}, \bar{X}^{(i-1)}$. So we draw

- $\mu^{(i)} | \phi^{(i-1)}, \theta^{(i-1)}, c_u^{(i-1)}, \sigma^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
- $\phi^{(i)} | \mu^{(i)}, \theta^{(i-1)}, c_u^{(i-1)}, \sigma^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
- $\theta^{(i)} | \mu^{(i)}, \phi^{(i)}, c_u^{(i-1)}, \sigma^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
- $c_u^{(i)} | \mu^{(i)}, \phi^{(i)}, \theta^{(i)}, \sigma^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
- $\sigma^{(i)} | \mu^{(i)}, \phi^{(i)}, \theta^{(i)}, c_u^{(i)} \epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
- $\epsilon_0^{(i)} | \mu^{(i)}, \phi^{(i)}, \theta^{(i)}, c_u^{(i)}, \sigma^{(i)} X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
- $X_0^{(i)} | \mu^{(i)}, \phi^{(i)}, \theta^{(i)}, c_u^{(i)}, \sigma^{(i)}, \epsilon_0^{(i)}, X^{(i-1)}, \bar{X}^{(i-1)}$

by using a *sequential slice sampler*. Here, we show how we draw $\theta^{(i)}$. The same analysis is repeated for all the above parameters in the given order.

(a) Since we always guarantee to have feasibility of $\bar{X}$ in each step, we can compute the likelihood

$$\begin{aligned} L_\theta^{(i-1)} &= L(\Lambda; X^{(i-1)}, \bar{X}^{(i-1)}) \\ &= L(\Lambda; X^{(i-1)}) \\ &= P(X_1^{(i-1)}, ..., X_T^{(i-1)} | \mu^{(i)}, \phi^{(i)}, \theta^{(i-1)}, c_u^{(i-1)} \epsilon_0^{(i-1)}, X_0^{(i-1)}) \end{aligned}$$

This is our vertical slice.

(b) Draw a random variable u_θ from Uniform$(0, L_\theta^{(i-1)})$. This is our horizontal slice.

(c) While true,

i. Draw $\theta^{(i)}$ from the prior distribution $\theta^{(i)} | \mu^{(i)}, \phi^{(i)}, c_u^{(i-1)}, \sigma^{(i-1)}$. We have

$$\Lambda' = \left| \begin{array}{c} \mu^{(i)} \\ \phi^{(i)} \\ \theta^{(i)} \\ \log(\sigma^{(i-1)}) \\ \log(c_u^{(i-1)}) \end{array} \right| \sim N(\hat{\Lambda}, V)$$

From proposition 1, we have the conditional truncated distribution

$$\theta^{(i)} | \mu^{(i)}, \phi^{(i)}, c_u^{(i-1)}, \sigma^{(i-1)} \sim N(\hat{\Lambda}_3 + V_{3,-3} V_{-3,-3}^{-1} (\Lambda'_{-3} - \hat{\Lambda}_{-3}), \hat{\Lambda}_{3,3} - \hat{\Lambda}_{3,-3} \hat{\Lambda}_{-3,-3}^{-1} \hat{\Lambda}_{-3,3})$$

for $-1 \leq \theta^{(i)} \leq 1$.

ii. Check the feasibility of the draw [1]. If the solution is not feasible go to Step (i) and make another draw for $\theta^{(i)}$. Otherwise, find the likelihood $L_\theta^{(i)} = P(X_1^{(i-1)}, ..., X_T^{(i-1)} | \mu^{(i)}, \phi^{(i)}, \theta^{(i-1)}, c_u^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)})$. If $L_\theta^{(i)} > u_\theta$, then keep $\theta^{(i)}$ and repeat the same analysis for the next parameter($c_u^{(i)}$ in this case).

2. In this step, we generate the $X^{(i)}, \bar{X}^{(i)} | \Lambda^{(i)}, \Upsilon^{(i)}, \hat{X}$. This can be done in two ways:

(a) We can draw $X_t | \Lambda^{(i)}, X_{-t}^{(i)}$ by using a sequential slice sampler for $t = 1, ..., T$ where

$$X_{-t}^{(i)} = \left\{ X_1^{(i)}, ..., X_{t-1}^{(i)}, X_{t+1}^{(i-1)}, ..., X_T^{(i-1)} \right\}$$

[1] For K=1, the parameters are always feasible. For $K > 1$ we need the feasibility which means that $X^{(i-1)}$ and $\Lambda^{(i)}$ generates the optimal forecast $\bar{X}$

i. We compute the likelihood $L_X^{(i-1)} = P(\gamma^{(i-1)}; \upsilon^{(i-1)})$ where

$$\gamma_t(k)^{(i-1)} = \hat{X}_t(k)^{(i-1)} - \bar{X}_t(k)^{(i-1)}$$

for $t = 1, ..., T$ and $k = 1, ..., K$.

This is our vertical slice.

ii. Draw a random variable u_X from Uniform$(0, L_X^{(i)})$. This is our horizontal slice.

iii. While true,

A. We can show that $X_t|\Lambda^{(i)}, X_{-t}^{(i)}$ follows a truncated normal distribution. Draw X_t from the truncated normal distribution.

B. Compute the likelihood $L_X^{(i)} = P(\gamma^{(i)}; \upsilon^{(i-1)})$. If $L_X^{(i)} > u_X$, then keep $X_t^{(i)}$ otherwise make another draw for X_t.

(b) This can also be done by drawing $\epsilon \sim N(0, \sigma^2 I)$ by using a slice sampler. In this case we draw ϵ together and compute the likelihood.

3. In this step, we generate $\Upsilon^{(i)}|\bar{X}^{(i)}, \hat{X}$. We first compute $\gamma^{(i)}$ as follows

$$\gamma_t(k)^{(i-1)} = \hat{X}_t(k)^{(i-1)} - \bar{X}_t(k)^{(i-1)}$$

for $t = 1, ..., T$ and $k = 1, ..., K$.

So we draw

- $p^{(i)}|\omega_1^{(i-1)}, \omega_2^{(i-1)}, \gamma^{(i)}$
- $\omega_1^{(i)}|p^{(i)}, \omega_2^{(i-1)}, \gamma^{(i)}$
- $\omega_2^{(i)}|p^{(i)}, \omega_1^{(i)}, \gamma^{(i)}$

by using a *sequential slice sampler*. Here, we show how we draw $p^{(i)}$. The same analysis is repeated for all the above parameters in the given order.

(a) We can compute the likelihood

$$L_p^{(i-1)} \quad = \quad P(\gamma^{(i)}|\Upsilon^{(i-1)})$$

This is our vertical slice.

(b) Draw a random variable u_p from Uniform$(0, L_p^{(i-1)})$. This is our horizontal slice.

(c) While true,

i. Draw $p^{(i)}$ from the prior $Beta(a, b)$.

ii. Find the likelihood $L_p^{(i)} = P(\gamma^{(i)}|p^{(i)}, \omega_1^{(i-1)}, \omega_2^{(i-1)})$. If $L_p^{(i-1)} > u_p$, then keep $p^{(i)}$ and repeat the same analysis for the next parameter($\omega_1^{(i)}$ in this case). Otherwise make another draw from the prior for $p^{(i)}$.

We also use shrinkage algorithm (Neal 2003) to improve the draws for the slice sampler. By doing that we decrease the number of rejected draws to find a feasible set of parameters.

4.3 Example

We use the customer forecasts in Example 1 to show the results of estimation. In Example 1, we only have forecasts for three weeks. By adding the forecasts for the following 17 weeks, we have the forecast vector $\hat{X}$ in Figure 4a. $\hat{X}$ provides $K = 5$ weeks forecasts in $T = 20$ weeks. We assume the following priors for estimation:

$$
\begin{array}{lll}
\mu \sim N(10, 5) & \phi \sim N(0, 0.5) & \theta \sim N(0, 0.5) \\
\sigma \sim IG(2.5, 10) & log(c_u) \sim N(0, 1) & \\
\epsilon \sim N(0, \sigma) & X_0 \sim N(\mu, \sigma\sqrt{\frac{1+\theta^2-2\theta\phi}{1-\phi^2}}) &
\end{array}
$$

We run our sampler for $40,000$ iterations to estimate the posterior distributions of each parameters. We discard the first $3,000$ observations as the warmup period. The results are available in Figure 4. Some of the inferences that we can make from the posterior distributions are as follows:

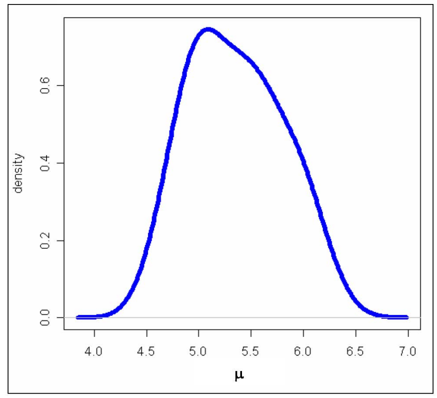
- We can see that the autoregressive parameter (ϕ) has a 95% confidence interval of $(0.59, 0.64)$. Therefore, there is a significant effect of the previous demand observation on the current demand. However, for θ, the posterior centers around 0. We cannot reject the hypothesis that $\theta = 0$. This means that error terms are not autocorrelated, however there is strong autocorrelation between the consecutive demand values.

- When we test the hypothesis that $c_u > c_o = 1$, we cannot reject it with $\alpha = 95\%$. This means that there is significant evidence from the data that the customer adds a positive bias to his forecasts.

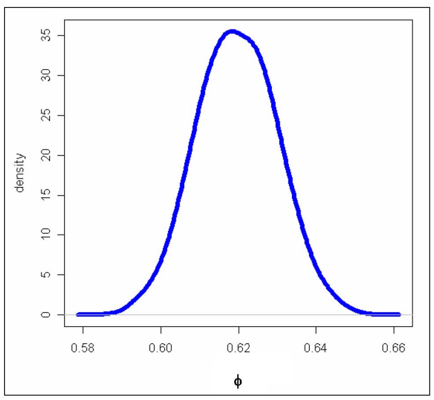
- When we look at σ, we can see that there is high uncertainty at the customer. The high variance is an important factor that causes errors in the forecasts. However, σ by itself does not explain the forecast errors. As we discuss above, the customer adds a bias to forecasts due to his asymmetric loss function.

$$\hat{X} = \begin{vmatrix} 5 & 5 & 6 & 6 & 6 \\ 1 & 3 & 4 & 5 & 6 \\ 2 & 4 & 5 & 5 & 6 \\ 6 & 6 & 6 & 6 & 6 \\ 5 & 6 & 6 & 6 & 6 \\ 7 & 7 & 6 & 6 & 6 \\ 7 & 7 & 7 & 7 & 7 \\ 4 & 5 & 5 & 6 & 6 \\ 4 & 5 & 5 & 6 & 6 \\ 3 & 4 & 5 & 6 & 6 \\ 5 & 6 & 6 & 6 & 6 \\ 4 & 5 & 6 & 6 & 6 \\ 3 & 5 & 5 & 6 & 6 \\ 6 & 6 & 6 & 6 & 6 \\ 7 & 7 & 7 & 7 & 6 \\ 10 & 8 & 8 & 7 & 7 \\ 8 & 7 & 7 & 7 & 7 \\ 9 & 8 & 8 & 7 & 7 \\ 8 & 7 & 7 & 7 & 7 \\ 4 & 5 & 6 & 6 & 6 \end{vmatrix}$$

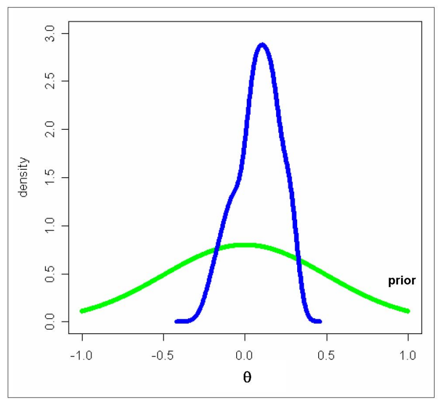
(a) $\hat{X}$



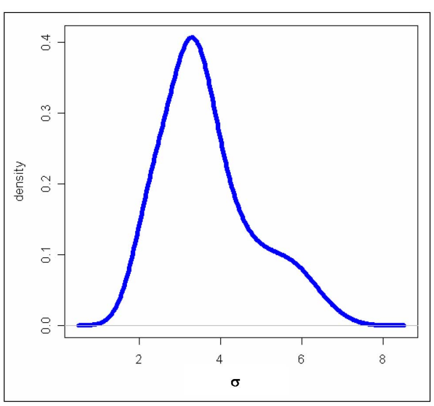
(b) 95% C.I.= $(4.61, 6.17)$



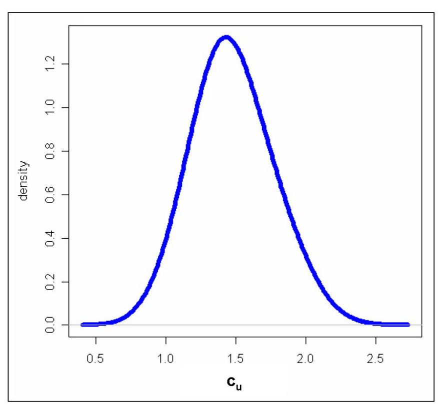
(c) 95% C.I.= $(0.59, 0.64)$



(d) 95% C.I.= $(-0.19, 0.29)$



(e) 95% C.I.= $(2.03, 6.31)$



(f) 95% C.I.= $(1.07, 1.96)$

Figure 4: The forecast vector $\hat{X}$ and the posterior distributions for μ, ϕ, θ, σ and c_u

The supplier can estimate the expected unbiased forecast by looking at the data. For example at $t = 1$, the expected unbiased forecasts for the next 5 periods is available in Figure 5. Therefore, by looking at a history of 20 periods with 5-step ahead forecasts, the supplier can have significant information about the forecast generation model of the customer and can remove the bias from the forecast.

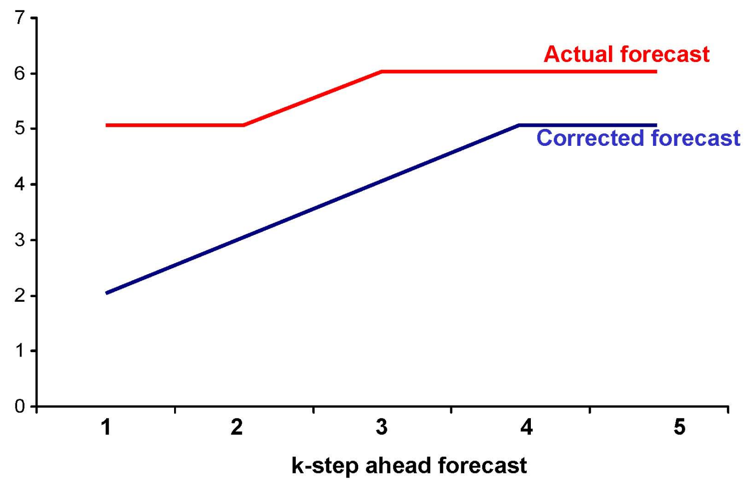


Figure 5: The actual and corrected forecast after the analysis at $t = 1$.

5 Hierarchical Model

In a hierarchical model, we study the diversity of the cost and demand parameters of the customers. Heterogeneity in cost and demand parameters give rise to different ordering behaviors from the customers. Depending on the diversity of the customers, hierarchical models provide better inferences about the forecast behavior of the customers. When the diversity is high, modeling the individual customer differences as well as the aggregate effects help the suppliers to better understand the forecast performance of the customers. In our analysis we have forecasts from multiple customers for multiple parts. Therefore, it is possible to make an inferences on the aggregate level as well as individual level. For example, we can analyze the forecasts of a customer for multiple parts and look at the forecast performance on an aggregate level. We can also perform an aggregate analysis for each part to understand the forecast performance of the manufacturers.

In our analysis we represent each customer with m and each part with n. We first define the first-stage priors as

$$\Lambda'_{mn} = \begin{vmatrix} \mu_{mn} \\ \phi_{mn} \\ \theta_{mn} \\ \log(\sigma)_{mn} \\ \log(c_u)_{mn} \end{vmatrix} \sim N(\hat{\Lambda}, V)$$

The second stage priors are

$$\begin{aligned} V &\sim \text{Inverted Wishart}(\nu_0, S_0) \\ \hat{\Lambda} &\sim N(0, \Lambda_0) \end{aligned}$$

In order to have a full rank $\nu_0 > 5$ which is the number of parameters. Assume that we run an analysis over M different customers for a part. So we can drop n from the subscript in this case, we have

$$P(V) \propto |V|^{-\frac{\nu_0-6}{2}} etr\left(-\frac{1}{2}S_0V^{-1}\right)$$

The posterior from the data is

$$\begin{aligned} P(V|\{\Lambda'_1, ..., \Lambda'_M\} &\propto |V|^{-\frac{M+\nu_0-6}{2}} etr\left(-\frac{1}{2}(S_0 + S+)V^{-1}\right) \\ &= \text{Inverted Wishart}(\nu_0 + M, S_0 + S) \end{aligned}$$

where $S = \sum_{m=1}^{M}(\Lambda'_m - \bar{\Lambda}')(\Lambda'_m - \bar{\Lambda}')^T$.

The posterior for $\hat{\Lambda}$ from the data is

$$\begin{aligned} P(\hat{\Lambda}|V, \{\Lambda'_1, ..., \Lambda'_M\} &\propto exp\left(\frac{1}{2}(\hat{\Lambda} - \hat{\Lambda}_M)^T V_M^{-1}(\hat{\Lambda} - \hat{\Lambda}_M)\right) \\ &= N(\hat{\Lambda}|\hat{\Lambda}_M, V_M) \end{aligned}$$

where

$$\begin{aligned} V_M &= \Lambda_0^{-1} + MV^{-1} \\ \hat{\Lambda}_M &= V_M^{-1}MV^{-1}\bar{\Lambda}' \end{aligned}$$

In order to predict the parameters, in iteration i we sequentially draw

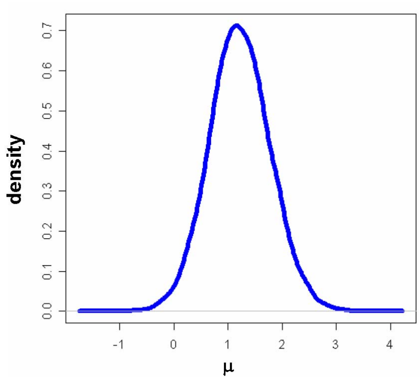
1. For customers $m = 1, ..., M$, we draw $\Lambda_m^{'(i)}|\Lambda^{'(i-1)}$ by using the sampler in section 4.2.
2. We draw $V^{(i)}|\{\Lambda_1^{'(i)}, ..., \Lambda_M^{'(i)}\}$ from the posterior distribution $N(\hat{\Lambda}|\hat{\Lambda}_M^{(i)}, V_M^{(i)})$.
3. We draw $\hat{\Lambda}^{(i)}|V^{(i)}, \{\Lambda_1^{'(i)}, ..., \Lambda_M^{'(i)}\}$ from the posterior distribution Inverted Wishart$(\nu_0 + M, S_0 + S^{(i)})$.

$$S_0 = \begin{bmatrix} 25 & 0 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 \\ 0 & 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 0 & 25 \end{bmatrix}$$

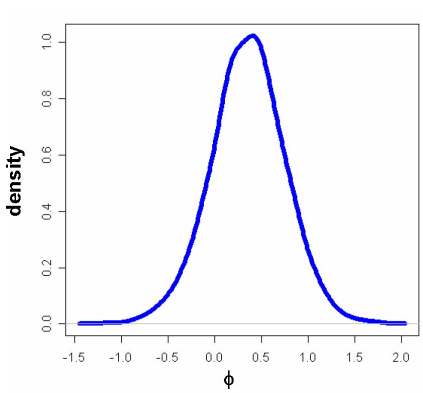
$$\nu_0 = 6$$

$$\Lambda_0 = \begin{bmatrix} 100 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 0 & 100 \end{bmatrix}$$

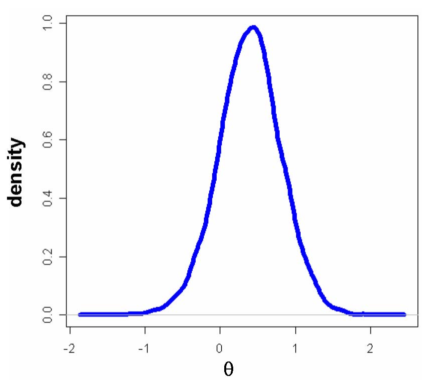
(a) The second stage priors



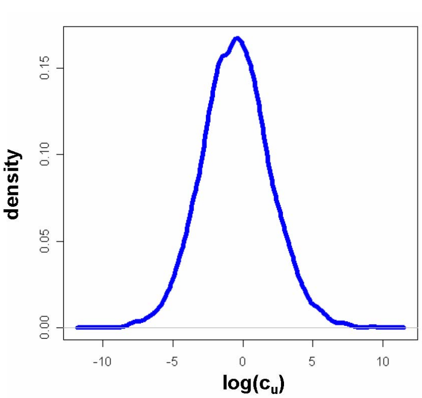
(b) Expected value for μ_i



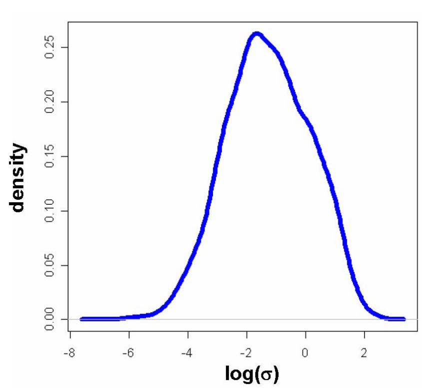
(c) Expected value for ϕ_i



(d) Expected value for θ_i



(e) Expected value for $\log(c_u)_i$



(f) Expected value for $\log(\sigma)_i$

Figure 6: The mean vector $\hat{\Lambda}$ for the hierarchical model with 16 customers and a single part

5.1 Example

In Figure 6, we provide the estimates for the mean level $\hat{\Lambda}$ of the model parameters. We analyze a part with 16 customers. During our analysis, we observe that for most of the customers, our model represents a good fit and we do not observe too many errors in the forecasts. For only 3 customers, we observe that the forecasts are generated from a different process. Therefore, the estimates for the model parameters were not significant for those customers.

From 6, we have the following observations:

1. We can see that ϕ and θ tends to be greater than zero. This means that the autocorrelations between the forecasts and the errors of the consecutive periods tend to be positive.

2. The log value of cost of underestimation is very close to normal distribution with a mean value slightly less than zero. This means that underestimation is also a common behavior among the customers. This also shows that log transformation is necessary while analyzing the cost of overestimation and underestimation. At the individual level, we get more significant results for the customers where we cannot reject the hypothesis for customers over- or underestimating their orders.

3. The standard deviation of the demand tends to be small for some of the customers. For those customers, we can argue that the forecast errors are mostly due to the cost structure of the customer rather than the uncertainty in usage.

6 Conclusion and Future Research

We have used a time-series framework to model the evolution of forecast vectors. In our data, we observe that customers consistently overestimate or underestimate their orders. The goal of our analysis was to develop a framework to test the significance of strategic behavior. By using asymmetric loss functions, we explain the dynamics behind the bias in customer's forecasts. We present an estimation procedure which has high predictive power to understand the cost and demand structure of a customer from his forecasts.

Our hierarchical model can be extended in order to include individual parameters for each customer and part. We can also include exogenous parameters in order to test their effect on the forecast performance of the customers. The hierarchical models for multiple customers and parents are ideal for gibbs sampling methods, where we sequentially draw individual parameters by conditioning on the other parameters. We mainly use the posterior distributions or slice sampling to draw the parameters. Other MCMC methods can also be considered for the analysis. The main advantage of the slice sampler is we do not need to tune the parameters and provide a proposal density.

We show that the customers add bias to their forecast due to their cost and demand structure. Therefore, it is critical for a supplier to create unbiased forecasts from the reported forecasts. Our analysis helps us understand which factors influence the forecast behavior of a manufacturer. A high σ for a customer shows that there is high uncertainty in the usage. A high c_u signals high cost of stockouts at the customer. We can also make inferences on the aggregate level. If we detect a common poor forecast performance across the customers, there might be problems with shipping of the product which creates artificial spikes in the forecasts. Hierarchical model helps us understand how much of the errors can be explained from the individual and aggregate demand and cost structure.

The customer demand can follow non-ARMA processes. In this case we observe high ω_1 or ω_2 values which show that the forecasts from the model do not fit the reported forecasts very well. Therefore, some other demand models can be considered for the non-ARMA customers. Our analysis can be extended to models with more components to capture these complexities. Possible such extensions are adding seasonality, production plans and nonlinear loss functions.

References

[1] N. Altintas and M. Trick. A data mining approach to forecast behavior of manufacturers in automotive industry. Working paper, Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, 2004.

[2] Y. Aviv. The effect of collaborative forecasting on supply chain performance. *Management Science*, 47:1326–1343, 2001.

[3] Y. Aviv. Gaining benefits from joint forecasting and replenishment processes: The case of auto-correlated demand. *Manufacturing and Service Operations Management*, 4:55–74, 2002.

[4] Y. Aviv. A time-series framework for supply-chain inventory management. *Operations Research*, 51(2):210–227, 2003.

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

[6] Roy Batchelor and Pami Dua. Forecaster diversity and the benefits of combining forecasts. *Management Science*, 41(1):68–75, 1995.

[7] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco, Holden-Day, 1970.

[8] G. Cachon and S. Netessine. Game theory in supply chain analysis. In S. D. Wu D. Simchi-Levi and Z. Shen, editors, *Handbook of Quantitative Supply Chain Analysis: Modeling in the eBusiness Era*. Kluwer, 2004.

[9] J. Geweke and C. Whiteman. Bayesian forecasting. In C.W.J. Granger G. Elliott and A. Timmermann, editors, *The Handbook of Economic Forecasting*. Elsevier, 2006.

[10] S. C. Graves. A tactical planning model for a job shop. *Operations Research*, 34:522–533, 1986a.

[11] S. C. Graves, D. B. Kletter, and W. B. Hetzel. A dynamic model for requirements planning with application to supply chain optimization. *Operations Research*, 46:S35–S49, 1998.

[12] S. C. Graves, D. B. Kletter, and W. B. Hetzel. A single-item inventory model for a non-stationary demand process. *Manufacturing and Service Operations Management*, 1(1):50–61, 1999.

[13] S. C. Graves, H.C. Meal, S. Dasu, and Y. Qiu. Two stage production planning in a dynamic environment. In S. Axsater, C. Schneeweiss, and E.Silver, editors, *Multi-Stage Production Planning and Control: Lecture Notes in Economics and Mathematical Systems*, pages 9–43. Springer-Verlag, Berlin, 1986b.

[14] R. Gullu. On the value of information in dynamic production/inventory problems under forecast evolution. *Naval Research Logistics*, 43:289–303, 1986.

[15] D. C. Heath and P. Jackson. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions*, 26:17–30, 1994.

[16] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistics*. Prentice Hall, 1992.

[17] M. A. Lariviere and E. L. Porteus. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science*, 45:346–363, 1999.

[18] R. M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–767, 2003.

[19] H. Scarf. Bayes solutions of the statistical inventory problem. *Annals of Mathematical Statistics*, 30:490–508, 1959.

[20] C. Terwiesch, J. Z. Ren, T. H. Ho, and M. Cohen. An empirical analysis of forecast sharing in the semiconductor equipment industry. Working paper, Wharton Business School, University of Pennsylvania, Philadelphia, PA, 2003.

[21] B. Toktay and L. M. Wein. Analysis of a forecasting-production-inventory system with stationary demand. *Management Science*, 47:1268–1281, 2001.

A Appendix

A.1 The conditional distribution of X_1

Proposition 1 *Let X be distributed as $N(\mu, \Sigma)$ and*

$$X = \begin{vmatrix} X_1 \\ X_2 \end{vmatrix} \quad \Sigma = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix} \quad \mu = \begin{vmatrix} \mu_1 \\ \mu_2 \end{vmatrix}$$

Then the conditional distribution of X_1, given X_2 is $N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

Proof: Available in Johnson and Wichern (1992). □

A.2 Matrix Form of the problem

Before we start looking at the supplier's problem, we introduce some concepts about multivariate normal distributions which helps to derive some useful results. We have X and $\hat{X}$ as follows

$$X = \begin{vmatrix} X_1 \\ X_2 \\ \vdots \\ X_{T-1} \\ X_T \end{vmatrix} \quad \hat{X} = \begin{vmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \vdots \\ \hat{X}_{T-1} \\ \hat{X}_T \end{vmatrix} = \begin{vmatrix} \hat{X}_1(1) & \hat{X}_1(2) & \cdots & \hat{X}_1(K-1) & \hat{X}_1(K) \\ \hat{X}_2(1) & \hat{X}_2(2) & \cdots & \hat{X}_2(K-1) & \hat{X}_2(K) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{X}_{T-1}(1) & \hat{X}_{T-1}(2) & \cdots & \hat{X}_{T-1}(K-1) & \hat{X}_{T-1}(K) \\ \hat{X}_T(1) & \hat{X}_T(2) & \cdots & \hat{X}_T(K-1) & \hat{X}_T(K) \end{vmatrix}$$

If we put the problem in matrix notation, we have the following model

$$\begin{aligned} (1-\phi B)X &= (1-\theta B)\epsilon + A_1 \\ X &= (1-\phi B)^{-1}(1-\theta B)\epsilon + (1-\phi B)^{-1}A_1 \end{aligned}$$

$$(1-\phi B) = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\phi & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -\phi & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\phi & 1 \end{vmatrix} \quad (1-\theta B) = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\theta & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\theta & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -\theta & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\theta & 1 \end{vmatrix}$$

$$A_1 = \begin{vmatrix} (1-\phi)\mu + \phi X_0 - \theta\epsilon_0 \\ (1-\phi)\mu \\ \vdots \\ (1-\phi)\mu \end{vmatrix} \quad (1-\phi B)^{-1} = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \phi & 1 & 0 & \cdots & 0 & 0 & 0 \\ \phi^2 & \phi & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \phi^{T-3} & \phi^{T-4} & \phi^{T-5} & \cdots & 1 & 0 & 0 \\ \phi^{T-2} & \phi^{T-3} & \phi^{T-4} & \cdots & \phi & 1 & 0 \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \cdots & \phi^2 & \phi & 1 \end{vmatrix}$$