A Theoretical Derivation of Relationships between Forecast Errors

Jerry Z. Shan

This paper studies errors in forecasting the demand for a component used by several products. Because data for the component demand (both actual demand and forecast demand) at the aggregate product level is easier to obtain than at the individual product level, the study focuses on the theoretical relationships between forecast errors at these two levels.

With a sound theoretical foundation for understanding forecast errors, a much more confident job can be done in forecasting and in related planning work, even under uncertain business conditions.[1]

In a typical material planning process, planners are constantly challenged by forecast inaccuracies or errors. For example, should a component forecast error be measured for each platform for which it may be needed, or should its forecast accuracy be measured at the aggregate level, across platforms? What is the relation between the two accuracy measures?

This paper describes a theoretical study of forecast errors. First, we formally define forecast errors with different rationales, derive several relationships among them, and prove a heuristic formula proposed by Mark Sower.[1] Then we study the effects of a systematic bias on the forecast errors. Finally, we extend our study to the situations where correlations across product demands and time effects in demand and forecast are taken into account. Definitions and theorems are presented first, and proofs of the theorems are given at the end of the paper.

Jerry Z. Shan

Jerry Shan is a software engineer at HP Laboratories, responsible for statistical analysis and experimental design for applications of HP Laboratories' Enterprise Modeling and Simulation system. He received his PhD degree in statistics from Stanford University and joined HP in 1994. Born in China's Jiangsu province, he taught statistics at China Textile University before coming to the Unites States. He is married, has two children, and enjoys soccer, photography, and swimming.

Basic Concepts

Consider the case of a component that can be used for the manufacture of n different products, or *platforms*. For platform i ($1 \leq i \leq n$), denote by F_i the forecast demand for the component, and by D_i the actual demand. In the treatment of forecast and actual, we propose in this paper the following framework: *Regard forecast demand as deterministic, or predetermined, and actual demand as stochastic.* By stochastic, we mean that given the same operating environment or experimental conditions, the actual demand can be different from one operation run to another. Thus, we can postulate a probability distribution for it.

For a generic case, denote by D the actual demand and by F the forecast. We call the forecast *unbiased* if $E(D) = F$, where $E(D)$ denotes the expectation, or expected value, of D with respect to its probability distribution. Practically speaking, this unbiased requirement means that over many runs under the same operating conditions, the average of the realized demand is the same as the forecast. If there is a deterministic quantity $b \neq 0$ such that $E(D) = F + b$, then we say the forecast is *biased*, and the bias is b. In practice, this means that there is a systematic departure of the average realized demand from the forecast.

Throughout the paper, we often make the normality assumption on the demand, that is, for unbiased forecasts, we assume that the demand D has a normal (Gaussian) distribution with mean F and standard deviation σ, that is, $D \sim N(F, \sigma^2)$. Is this a reasonable assumption in reality? The answer is yes. First of all, this assumption is technically equivalent to assuming that the difference $\varepsilon = D - F$ between the actual demand D and the forecast F is normally distributed: $\varepsilon \sim N(0, \sigma^2)$. The validity of this latter assumption is based on the fact that the difference between the actual demand and a good forecast is some aggregation of many small random errors, and on the central limit theorem, which states that the aggregation of many small random errors has a limiting normal distribution.

Unbiased Forecast Case

In this section, we assume unbiased forecasting at all platforms.[1] Statistically, $E(D_i) = F_i$, where F_i is the forecast for the common component at platform i, and D_i is the actual demand of the component at platform i.

Definition 1: (Same Weight Mean Based) Define $E_\pi = E(\varepsilon_\pi)$ to be the forecast error at the mean (average) platform level, and $E_a = E(\varepsilon_a)$ to be the forecast error at the aggregate platform level, where:

$$\varepsilon_\pi = \frac{1}{n}\sum_{i=1}^{n} \frac{|D_i - F_i|}{F_i}, \tag{1a}$$

and

$$\varepsilon_a = \frac{\left|\sum_{i=1}^{n} D_i - \sum_{i=1}^{n} F_i\right|}{\sum_{i=1}^{n} F_i}. \tag{1b}$$

The rationale of defining the forecast error at the mean level and at the aggregate level is as follows. Let $\varepsilon_i = |D_i - F_i|/F_i$. Then ε_i measures, in terms of the relative difference, the forecast error at a single platform i. Accordingly, ε_a measures the forecast error, also in terms of the relative difference, at the aggregate level from all platforms, and ε_π provides an estimate for the forecast error at any individual platform since it is the average of the forecast errors over all individual platforms. Because all the quantities in equation 1 are stochastic, we take expectations to get their deterministic means. Now, a natural question is: What is the relation between the errors at the two different levels?

Theorem 1: Based on definition 1, and assuming that $D_i \sim N(F_i, \sigma^2)$, i = 1, 2, ..., n, and that the D_i are uncorrelated (strictly speaking, we also need the joint normality assumption, which in general can be satisfied), we have:

1. $E_\pi = \sqrt{n} E_a C_n$, where:

$$C_n = \left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{F_i}\right)\left(\frac{1}{n}\sum_{i=1}^{n} F_i\right). \tag{2}$$

2. It is always true that $C_n \geq 1$, and $C_n = 1$ if and only if the forecasts across all the platforms are the same.

We note that in the definition for ε_π, we used the same weight, 1/n, for all platforms. If instead we use a weight proportional to the forecast at the platform, then we have the following:

Definition 2: (Weighted Mean Based) Define $E_\pi = E(\varepsilon_\pi)$ and $E_a = E(\varepsilon_a)$, where:

$$\varepsilon_\pi = \sum_{i=1}^{n} \frac{F_i}{\sum_{j=1}^{n} F_j} \frac{|D_i - F_i|}{F_i} = \frac{\sum_{i=1}^{n} |D_i - F_i|}{\sum_{i=1}^{n} F_i} \tag{3a}$$

and

$$\varepsilon_a = \frac{\left|\sum_{i=1}^{n} D_i - \sum_{i=1}^{n} F_i\right|}{\sum_{i=1}^{n} F_i}. \quad (3b)$$

Theorem 2: Based on definition 2 and with the same assumptions as in theorem 1, we have:

$$E_\pi = \sqrt{n}\, E_a. \quad (4)$$

Mark Sower[1] proposed this heuristic formula. Theorem 2 says that under suitable conditions, equation 4 holds exactly.

Other researchers have addressed a similar problem from the perspective of demand variability. In measuring the relative errors of the forecast at the individual platforms, it was assumed that σ_i/μ_i $(i = 1, 2, ..., n)$ are the same, where σ_i is a measure of demand variability and μ_i is the mean demand at platform i. The advantage here is we do not need to make such a strong assumption. In fact, our measure of the forecast error at the individual platform level can be interpreted as the forecast error at an averaged individual platform.

The following definition of error is based on this observation in practice. The standard deviation of a random variable can be very large if the values this random variable takes on are very large. A more sensible error measure of such a random variable would be the relative error rather than the absolute error. So, given a random variable X, we can measure its error by the coefficient of variance $cv(X) = \sigma(X)/E(X)$ rather than by its standard deviation $\sigma(X)$.

With the unbiased forecast assumption, the forecast error at platform i can be measured by $cv(D_i)$. The average of these coefficients over all platforms is a good measure of the forecast error at the individual platform level. On the other hand, $\sum_{i=1}^{n} D_i$ is the demand from all platforms, and $\sum_{i=1}^{n} F_i$ is the corresponding forecast, so $cv\left(\sum_{i=1}^{n} D_i\right)$ is a good measure of the forecast error at the aggregated platform level.

Definition 3: (CV Based) Define:

$$E_\pi = \sum_{i=1}^{n} cv(D_i)/n \text{ and } E_a = cv\left(\sum_{i=1}^{n} D_i\right).$$

Theorem 3: Based on definition 3, and assuming that the D_i are uncorrelated, we have

$$E_\pi = \sqrt{n}\, E_a C_n, \quad (5)$$

where C_n is defined in equation 2. For theorem 3, we do not have to assume normality to get the relevant results. This is also true for theorem 4.

General Case: The Effect of Bias

We assume here that forecasts are consistently biased. This is expressed as $E(D_i) = F_i + b$, where b denotes the common forecast bias. This indicates that F_i overestimates demand when $b < 0$ and underestimates demand when $b > 0$.

Can we extend the use of definition 3 for the forecast errors to this general case? The answer is no. This is because the standard deviation is independent of bias, and therefore one could erroneously conclude that the forecast error is small when the standard deviation is small, even though the bias b is very significant. Instead, the forecast error now should be measured by the functional:

$$e(D, F) = \sqrt{E([D - F]^2)}\,/F, \quad (6)$$

rather than by the cv, which is $\sqrt{E([D - E(D)]^2)}\,/E(D)$. Hence, in parallel with definition 3, we have the following definition.

Definition 4: (e-Functional Based) Define:

$$E_c = e\left(\sum_{i=1}^{n} D_i, \sum_{i=1}^{n} F_i\right) \text{ and } E_\pi = \sum_{i=1}^{n} e(D_i, F_i)/n,$$

where the functional e is defined in equation 6.

If the bias $b = 0$, then the functional e in equation 6 is the same as the cv, and hence definitions 3 and 4 are equivalent.

Theorem 4: Based on definition 4, and assuming that $D_i \sim (F_i + b, \sigma^2)$,* $i = 1, 2, ..., n$ and that the D_i are uncorrelated, we then have:

$$E_\pi = \sqrt{n}\, E_a C_n \sqrt{\frac{\sigma^2 + b^2}{\sigma^2 + nb^2}}, \qquad (7)$$

where C_n is given in equation 2.

Since definition 1 considers the relative difference between the forecast and the actual, any bias in the forecast will be retained in the difference, so there is no problem in using this definition even if there is bias. However, the relation between the two errors has changed.

Theorem 5: Based on definition 1 and the assumption that $D_i \sim N(F_i + b, \sigma^2)$, $i = 1, 2, ..., n$ and that the D_i are uncorrelated, we have:

$$E_\pi = \sqrt{n}\, E_a C_n \frac{\sqrt{\frac{2}{\pi}}\sigma e^{-b^2/2\sigma^2} + b[2\Phi(b/\sigma) - 1]}{\sqrt{\frac{2}{\pi}}\sigma e^{-b^2/2\sigma^2} + \sqrt{n}\, b[2\Phi(\sqrt{n}\, b/\sigma) - 1]}, \qquad (8)$$

where C_n is defined in equation 2, and $\Phi(x)$ is the cumulative distribution function of the standard normal distribution $N(0, 1)$ at x.

If there is no bias in the forecasting, the relationships between the errors at the two levels are exactly the same for definitions 1 and 3: $E_\pi = \sqrt{n}\, E_a C_n$. This formula, with the introduction of the constant C_n, is slightly different from the hypothesized equation 4. As noted in theorem 1, it is always true that $C_n \geq 1$. If we use definition 2, then equation 4 holds exactly.

If there is bias in the forecasting, then in each relationship formula (equation 7 or equation 8), there is another multiplying factor that reflects the effect of the bias. One can easily find that both of these multiplying factors are less than or equal to 1. This implies that, compared to the error at the component level, the error at the platform-component level when forecast bias exists is less than when the forecast bias does not exist.

If bias does exist, as it does in reality, it seems that the multiplying factor resulting from bias in either equation 7 or equation 8 should be taken into consideration, with suitable estimation of the parameters involved.

* The notation $X \sim (\mu, \sigma^2)$ means that X has mean μ and standard deviation σ but is not necessarily normally distributed.

Correlated Demands

It is reasonable to assume that demand for a component for one platform affects demand for this component for another platform. Also, for a given platform, there is usually a strong correlation between the current demand and the historical demands. The forecast is usually made based on the historical demands. In this section, we first propose a correlated multivariate normal distribution model for the demand stream when the platform is indexed, and then propose a time-series model for the demand and forecast streams when time is indexed. Our goal is to expand our study of the relationship between the two layers of forecast errors in the presence of correlations. Throughout this section, we assume unbiased forecasts, and use the weighted average definition (definition 3) for the forecast error.

Correlated Normal Distribution Model at a Time Point. In this subsection we consider the case where there is correlation across platform demands, but we still assume that time does not affect demand. Suppose that the demand stream D_i, $i = 1, 2, ..., n$ can be modeled by a correlated normal distribution such that $D_i \sim N(F_i, \sigma^2)$ for $i = 1, 2, ..., n$ and that there is a correlation between different D_i expressed as $Cov(D_i, D_j) = \sigma^2 \rho_{ij}$ for $1 \leq i \neq j \leq n$. With this assumption on the demand stream, we have the following result.

Theorem 6: Based on definition 2 and the above correlated normal distribution modeling for the demand stream, we have:

$$E_\pi = \frac{n}{\sqrt{\sum_{1 \leq i \neq j \leq n} \rho_{ij} + n}} E_a. \qquad (9)$$

In particular, if $\rho_{ij} = \rho$ for all $1 \leq i \neq j \leq n$, then we get:

$$E_\pi = \frac{\sqrt{n}}{\sqrt{(n-1)\rho + 1}} E_a. \qquad (10)$$

When the common correlation coefficient ρ is 0 or near 0, we see that equation 4 holds exactly or approximately.

Autoregressive Time Series Model. Now we take into consideration the time effect in the product demand. For platform i, $i = 1, 2, ..., n$ at time t, $t = 1, 2, ...,$ denote by $D_i^{(t)}$

the demand and $F_i^{(t)}$ the forecast. Suppose that the demand stream over time at each platform can be modeled by an autoregressive model AR(p). At platform i, the autoregressive model assumes that the demand at the current time t is a linear function of the past demands plus a random disturbance, that is:

$$D_i^{(t)} = \sum_{j=1}^{p} a_{i,j} D_i^{(t-j)} + \varepsilon_i^{(t)},$$

where the $a_{i,j}$ are constant coefficients. Further, suppose that the forecast $F_i^{(t)}$ is optimal given the historical demand profile $\mathcal{F}_i^{(t-1)} = \sigma\left(D_i^{(1)},\ D_i^{(2)},\ \ldots,\ D_i^{(t-1)}\right)$. That is, with $D_i^{(0)},\ D_i^{(-1)},\ \ldots,\ D_i^{(-(p-1))}$ properly initialized, for $t \geq 1$:

$$D_i^{(t)} = a_{i,1} D_i^{(t-1)} + a_{i,2} D_i^{(t-2)} + \cdots + a_{i,p} D_i^{(t-p)} + \varepsilon_i^{(t)},$$

and

$$F_i^{(t)} = E\left(D_i^{(t)} \middle| \mathcal{F}_i^{(t-1)}\right)$$
$$= a_{i,1} D_i^{(t-1)} + a_{i,2} D_i^{(t-2)} + \cdots + a_{i,p} D_i^{(t-p)},$$

where $\varepsilon_i^{(1)},\ \varepsilon_i^{(2)},\ \ldots,\ \varepsilon_i^{(t)},\ \ldots$ are independently and identically distributed as $N(0, \sigma^2)$ and the random disturbance at time t, that is, $\varepsilon_i^{(t)}$, is independent of the demand stream before time t, that is, $\{D_i^{(t-1)},\ D_i^{(t-2)},\ \ldots\}$. Also, we assume independence across platforms. With the above modeling of the demand and forecast, what can we say about the relationship between the two layers of forecast errors?

Theorem 7: Based on definition 1 or 2 and the above time-series modeling for the demand stream and forecast stream, and assuming that the variances at all platforms are the same, then at any time point, if definition 1 is used:

$$E_\pi^{(t)} = \sqrt{n}\, E_a^{(t)} \tilde{C}_n, \tag{11}$$

where

$$\tilde{C}_n = \frac{E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{F_i^{(t)}}\right)}{E\left[\frac{n}{\sum_{i=1}^{n} F_i^{(t)}}\right]}.$$

and if definition 2 is used, then:

$$E_\pi^{(t)} = \sqrt{n}\, E_a^{(t)}. \tag{12}$$

Rewriting C_n in equation 2 as

$$C_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{F_i^{(t)}}}{\frac{n}{\sum_{i=1}^{n} F_i^{(t)}}}$$

and taking expectations for the numerator and denominator separately in the expression leads to $\tilde{C}_n$. Hence, it is always true that $\tilde{C}_n \geq 1$.

Proofs

Theorem 1 is a special case of theorem 5. Theorem 3 is a special case of theorem 4. The proof for theorem 6 is similar to that for theorem 5, with an application of lemma 1.

Lemma 1: If $X \sim N(b, \sigma^2)$, then:

$$E|X| = \sqrt{\frac{2}{\pi}}\sigma e^{-b^2/2\sigma^2} + b[2\Phi(b/\sigma) - 1] \equiv H(b, \sigma). \tag{13}$$

Proof of Lemma 1: Without loss of generality, we can assume that $\sigma = 1$, since otherwise we can make a simple transformation $Y = X/\sigma$.

$$E|X| = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} |x| e^{-(x-b)^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} |x| e^{-(x-b)^2/2} dx + \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} |y| e^{-(y+b)^2/2} dy$$

$$= I(b) + I(-b), \text{ where}$$

$$I(b) \equiv \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-(x-b)^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-b}^{\infty} (y + b) e^{-y^2/2} dy$$

$$= \frac{1}{\sqrt{2\pi}} e^{-b^2/2} + b\Phi(b), \text{ and hence}$$

$$E|X| = \frac{2}{\sqrt{2\pi}} e^{-b^2/2} + b\Phi(b) + (-b)\Phi(-b)$$

$$= \sqrt{\frac{2}{\pi}} e^{-b^2/2} + b[2\Phi(b) - 1].$$

Proof of Theorem 1 Parts 2 and 3. First note that function $\varphi(x) = 1/x$ is convex over $(0, \infty)$. Let random variable X have a uniform distribution on the set $\{F_i: 1 \le i \le n\}$, that is, $P(X = F_i) = 1/n$. An application of the Jensen inequality[2] $E\varphi(X) \ge \varphi(EX)$ leads to the desired inequality. The second part is based on the condition for the Jensen inequality to become an equality.

Proof of Theorem 4:

$$e(D_i, F_i) = \frac{\sqrt{E([D_i - F_i]^2)}}{F_i} = \frac{\sqrt{\sigma^2 + b^2}}{F_i},$$

$$E_\pi = \frac{1}{n}\sum_{i=1}^{n} e(D_i, F_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{\sigma^2 + b^2}}{F_i},$$

$$E_a = e\left(\sum_{i=1}^{n} D_i, \sum_{i=1}^{n} F_i\right) = \frac{\sqrt{n\sigma^2 + (nb)^2}}{\sum_{i=1}^{n} F_i}.$$

Hence we have:

$$\frac{E_\pi}{E_a} = \sqrt{n}\sqrt{\frac{\sigma^2 + b^2}{\sigma^2 + nb^2}} C_n.$$

Proof of Theorem 5: Noting that:

$$D_i - F_i \sim N(b, \sigma^2) \text{ and } \sum_{i=1}^{n}(D_i - F_i) \sim N(nb, n\sigma^2),$$

then we have:

$$\frac{E_\pi}{E_a} = \frac{E(e_\pi)}{E(e_a)}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{F_i}}{\frac{1}{\sum_{i=1}^{n} F_i}} \frac{H(b, \sigma)}{H(nb, \sqrt{n}\sigma)} \text{ (by lemma 1)}$$

$$= \sqrt{n} C_n \frac{\sqrt{\frac{2}{\pi}}\sigma e^{-b^2/2\sigma^2} + b[2\Phi(b/\sigma) - 1]}{\sqrt{\frac{2}{\pi}}\sigma e^{-b^2/2\sigma^2} + \sqrt{n}\, b[2\Phi(\sqrt{n}\, b/\sigma) - 1]}.$$

Proof of Theorem 7: The proofs for equations 11 and 12 are similar. We give a proof for equation 11 only. First notice that $D_i^{(t)} - F_i^{(t)} = \varepsilon_i^{(t)} \sim N(0, \sigma_i^2)$. At any given time t, by the definitions for $E_\pi^{(t)}$ and $E_a^{(t)}$, we have:

$$E_\pi^{(t)} = \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\left|\varepsilon_i^{(t)}\right|}{F_i^{(t)}}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left(\left|\varepsilon_i^{(t)}\right|\right) E\left(\frac{1}{F_i^{(t)}}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \sqrt{\frac{2}{\pi}}\sigma E\left(\frac{1}{F_i^{(t)}}\right).$$

This second step follows from the fact that $\varepsilon_i^{(t)}$ is independent of demands before time t, and hence independent of the optimal forecast at time t, $F_i^{(t)}$. The last step follows from lemma 1 and the same variance assumption across platforms.

$$E_a^{(t)} = E\left[\frac{\left|\sum_{i=1}^{n}\varepsilon_i^{(t)}\right|}{\sum_{i=1}^{n} F_i^{(t)}}\right]$$

$$= E\left(\left|\sum_{i=1}^{n}\varepsilon_i^{(t)}\right|\right) E\left[\frac{1}{\sum_{i=1}^{n} F_i^{(t)}}\right]$$

$$= \sqrt{\frac{2}{\pi}\sum_{i=1}^{n}\sigma_i^2}\, E\left[\frac{1}{\sum_{i=1}^{n} F_i^{(t)}}\right]$$

$$= \sqrt{n}\,\sigma\sqrt{\frac{2}{\pi}}\,\sigma E\left[\frac{1}{\sum_{i=1}^{n} F_i^{(t)}}\right].$$

The reasoning is the same as for proving $E_\pi^{(t)}$ above.

Conclusion

Forecast errors increase the complexity and difficulty of the production planning process. This results in excessive inventory costs and reduces on-time delivery. In this paper we have studied the forecast errors for the case of several products using the same component. Because data for the component demand (both actual demand and forecast demand) is easier to obtain at the aggregate product level than at the individual product level, we focused on the theoretical relationships between forecast errors at these two levels.

Our first task was to propose formal definitions for measuring forecast errors under different rationales and technical assumptions. The second task was to formally derive relationships between forecast errors at the two levels. As part of our work we proved the validity of a heuristic formula proposed by Mark Sower of the business operations planning department at the HP Roseville, California site.

In addition to analyzing the two-level problem, we derived a theoretical basis for relaxing the usual assumptions concerning correlations in the data across products and over time.

Acknowledgments

I offer my first thanks to Pano Santos for introducing me to a related material planning problem, and to Mark Sower for his marvelous intuition. Special thanks go to Farid Aitsahlia for his careful technical reading and editorial assistance. Thanks also go to Shahid Mujtaba, Alex Zhang (University of Southern California), and Dirk Beyer for their valuable comments and suggestions, and to Bob Ritter, Shailendra Jain, and Paul Williams for their management support and encouragement. In particular, Paul Williams helped greatly in writing the conclusion.

References

1. P. Santos, J. Shan, M. Sower, and A. Zhang, *Material Planning in Configure-To-Order Environment with Product Demand Uncertainty and Component Shortage Conditions*, HP Laboratories Technical Report HPL-97-32, 1997 (HP Internal Only).

2. E.B. Manoukian, *Modern Concepts and Theorems of Mathematical Statistics*, 1986.

▶ Go to Next Article

▶ Go to Journal Home Page