

Estimating Demand Uncertainty Using Judgmental Forecasts^{*}

Vishal Gaur[†], Saravanan Kesavan[‡], Ananth Raman[§], Marshall L. Fisher^{**}

April 15, 2005

Abstract

Measuring demand uncertainty is a key activity in supply chain planning. Of various methods of estimating the standard deviation of demand, one that has been employed successfully in the recent literature uses dispersion among experts' forecasts. However, there has been limited empirical validation of this methodology. In this paper we provide a general methodology for estimating the standard deviation of a random variable using dispersion among experts' forecasts. We test this methodology using three datasets, demand data at item level, sales data at firm level for retailers, and sales data at firm level for manufacturers. We show that the standard deviation of a random variable (demand and sales for our datasets) is positively correlated with dispersion among experts' forecasts. Further we use longitudinal datasets with sales forecasts made 3-9 months before earnings report date for retailers and manufacturers to show that the effects of dispersion and scale on standard deviation of forecast error are consistent over time.

* The authors gratefully acknowledge James Zeitler, Research Database Analyst at Harvard Business School for help in obtaining and preparing the data for this study.

[†] Leonard N. Stern School of Business, New York University, 44 West 4-th St., New York, NY 10012, Ph: 212-998-0297, Fax: 212-995-4227. E-mail: vgaur@stern.nyu.edu.

[‡] Harvard Business School, T-81C, Morgan Hall, Soldiers Field, Boston, MA 02163, Ph: 617 496-9890, Fax: 617 496-4059, E-mail: skesavan@hbs.edu.

[§] Harvard Business School, T-11, Morgan Hall, Soldiers Field, Boston, MA 02163, Ph: 617 495-6937, Fax: 617 496-4059, E-mail: araman@hbs.edu.

^{**} The Wharton School, University of Pennsylvania, Jon M. Huntsman Hall, 3730 Walnut St., Philadelphia, PA 19104-6366, Ph: 215 898-7721, Fax: 215 898-3664, E-mail: fisher@wharton.upenn.edu.

Estimating Demand Uncertainty Using Judgmental Forecasts

Abstract

Measuring demand uncertainty is a key activity in supply chain planning. Of various methods of estimating the standard deviation of demand, one that has been employed successfully in the recent literature uses dispersion among experts' forecasts. However, there has been limited empirical validation of this methodology. In this paper we provide a general methodology for estimating the standard deviation of a random variable using dispersion among experts' forecasts. We test this methodology using three datasets, demand data at item level, sales data at firm level for retailers, and sales data at firm level for manufacturers. We show that the standard deviation of a random variable (demand and sales for our datasets) is positively correlated with dispersion among experts' forecasts. Further we use longitudinal datasets with sales forecasts made 3-9 months before earnings report date for retailers and manufacturers to show that the effects of dispersion and scale on standard deviation of forecast error are consistent over time.

1. Introduction

Measuring demand uncertainty is a key activity in supply chain planning. Of various methods of estimating the standard deviation of demand, one that has been employed successfully in the recent literature uses dispersion among experts' forecasts. Experts on a panel independently assign point forecasts of demand for a given product and the standard deviation of these forecasts is used to estimate the standard deviation of demand. Thus far, however, there has been limited empirical validation of this methodology.

In this paper, we test the hypothesis that dispersion among experts' forecasts is positively correlated with the standard deviation of a random variable, and provide a general methodology for using dispersion among experts' forecasts to estimate the standard deviation of the random

variable. We define the standard deviation of a random variable as the standard deviation of its forecast error given forecasts by several experts. The principal challenge in estimating the foregoing relationship is that the standard deviation of the forecast error is not known and must be estimated from only a single realization of the random variable. We first propose a multiplicative functional form to represent the relationship between standard deviation of a random variable and dispersion among experts. We then derive a heteroscedastic regression equation to test the hypothesis and estimate the standard deviation of forecast error as a function of the dispersion among experts' point forecasts and other control variables. Heteroscedastic regression models have been well-studied in the econometrics literature. Our analysis applies the results of Harvey (1976).

We apply this methodology to three datasets. The first, obtained from Sport Obermeyer, an Aspen, Colorado-based ski-wear manufacturer, records demand forecasts and actual demand for 248 items at the style-color level. The second dataset, obtained from I/B/E/S, records at firm-level experts' forecasts of annual sales and realized annual sales for public-listed U.S. retailers for the years 1997-2004. We consider forecasts made at least 90 days prior to the announcement of actual sales, and partition our data into seven time buckets depending on the time difference between the forecast date and the earnings report date (when sales are revealed). We estimate our model for each of these time buckets, and thus, test the correlation between dispersion among experts and standard deviation of forecast error over different time frames. The third dataset, also obtained from I/B/E/S, pertains to public-listed U.S. manufacturers, and is analyzed similar to the second dataset.

Our empirical results show that the standard deviation of forecast error has a statistically significant positive correlation with dispersion among experts for all datasets. Testing this correlation using both non-parametric and parametric methods yields consistent results. We control for bias in forecasts as well as the effect of scale on the standard deviation of forecast error. As expected, scale is found to be statistically significant. Dispersion among experts'

forecasts has a statistically significant positive correlation with standard deviation of forecast error even after controlling for scale. We obtain fifteen estimates of the coefficients of $\log(\text{dispersion})$ and $\log(\text{scale})$ across the three datasets and different time buckets. The average values of the coefficients are 0.201 and 0.715, respectively.

Standard deviation is used extensively in both forecasting and inventory management. In the forecasting process the standard deviation of forecast errors is used to generate prediction intervals for the forecasts (Makridakis and Wheelwright 1987), in inventory management to compute order quantities. In many cases the standard deviation of forecast errors can be computed from historical forecasts and actual demand. But the method cannot be used for new products that lack demand history. Our methodology is particularly valuable in such cases. It can also be integrated with a causal forecasting model by including explanatory variables other than experts' forecasts.

Given the widespread use of judgemental forecasts in practice, there is a rich literature in experimental psychology on eliciting the standard deviation of a random variable from experts. Soll and Klayman (2004), for example, show experts to be grossly overconfident when asked to estimate the confidence interval of a random variable. Our work augments this literature in that our methodology requires experts to provide only point estimates of mean, while the papers in experimental psychology have sought to elicit confidence intervals from experts.

A number of contemporaneous research papers suggest a relationship between level of experts' agreement and degree of uncertainty of the variable they are trying to predict. The Delphi system of forecasting, for example, is used widely in operations management and other fields to obtain consensus opinions by reducing divergence among experts' views. Dalkey and Helmer (1963), note that the Delphi method often leaves some residual disagreement, which they interpret as a measure of the uncertainty of the variable being forecasted. Similarly, in numerous papers that discuss implementations of Delphi methodology it is stated that it was not possible to

obtain a consensus opinion due to various uncertainties faced by the forecasters (Milkovich et al. 1972; Munier and Ronde 2001).

A more direct relationship than Dalkey and Helmer (1963) was proposed by Fisher and Raman (1996). They found at Sport Obermeyer that the standard deviation of forecast error is highly correlated with the level of agreement among the experts who forecast demand for a product. Similarly, MacCormack and Verganti (2003), who used data collected from a Delphi process to assess level of market uncertainty, took the residual variation in experts' opinions that is, the variation that cannot be removed through the Delphi process, to be a measure of uncertainty around customers' requirements for the product.

As in operations management, experts, referred to as analysts in the finance literature, are employed to forecast various financial measures for firms. These analysts, however, are usually associated with the financial intermediaries that closely track firms' performance. Numerous articles that explore the characteristics of analysts' forecasts have been published. Givoly and Lakonishok (1984) survey the finance literature to summarize the findings on the properties of financial analysts' earnings forecasts. Malkiel (1982) and Ajinkya and Gift (1985) state that dispersion among analysts' earnings forecasts is a better measure of risk than beta. Whereas Ajinkya and Gift (1985) test the relationship between dispersion among analysts' forecasts and *implied* standard deviation (obtained from the Black-Scholes formula), we develop a model to test the relationship directly by estimating standard deviation in an operational context.

The rest of the paper is organized as follows. The main hypothesis of the paper is presented in §2. Section 3 deals with the model that is used for empirical analysis. The description of the data is presented in §4. Results are discussed in §5. Section 6 concludes this paper.

2. Hypothesis

The uncertainty of a random variable is often the result of many complex underlying processes that cannot be completely identified. As a result, experts tend to use different information sets or

forecasting models that capture different aspects of the information available for prediction. Hence experts could offer different forecasts for the same random variable.

We hypothesize that the extent to which experts differ from each other depends on the uncertainty of the random variable that they try to predict. In other words, we expect the dispersion among the experts' forecasts to be correlated with the standard deviation of the random variable. For example, suppose that the demand for a product is dependent on several factors, such as consumer tastes, weather patterns, fluctuations in the economy, alterations to the product as it evolves from design to production, etc. If some of these factors are uncertain, and experts have different views about them or each expert evaluates only a subset of the factors, then the forecasts given by experts will differ from each other. This reasoning is consistent with the research literature in psychology and decision sciences. Experts are known to differ in their forecasts of a random variable. In fact, Hogarth (1978) suggests that the experts that are not strongly correlated with each other should be selected for forecasting, and the optimal number of experts that a decision-maker should consult for forecasting is 8-12. Other researchers have sought to explain why experts differ in their forecasts. Summers, et al (2004) discuss how experts differ in terms of their functioning and experience; Armstrong (1986) states that experts differ when they make adjustments for recent events whose effects have not yet been observed; Flugstad and Windschitl (2003) show that even numerical estimates of probability can be interpreted in different ways depending on the reasons provided for arriving at the estimate.

Based on demand data from Sport Obermeyer we find that experts tend to be more accurate when they agree than when they do not agree. Figure 1 illustrates the relationship between forecast error and dispersion among expert's forecasts. Here, the *forecast error of demand* is defined as the absolute value of difference between the realized value of demand and the experts' mean forecast, and *dispersion among experts' forecasts* is defined as the standard deviation of the point forecasts of the different experts. For clarity, data points have been combined into groups of four. (The data are first sorted on dispersion among experts' forecasts

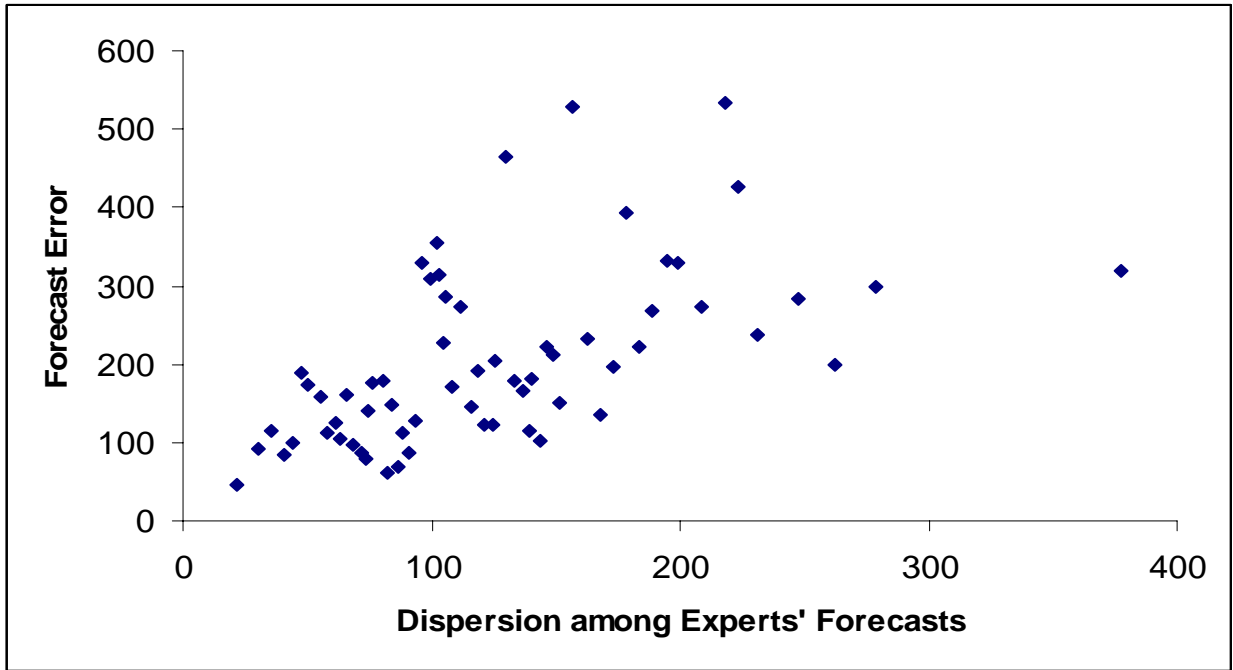


Figure 1: Relationship between Forecast Error and Dispersion among Experts' Forecasts at Sport Obermeyer.

and then combined into groups of four.) The figure shows a positive correlation between dispersion and forecast error. Motivated by this observation, we set up the following hypothesis:

Hypothesis 1. *Standard deviation of forecast error is positively correlated with the dispersion among experts' forecasts.*

We test this hypothesis against the null hypothesis that the standard deviation of forecast error is uncorrelated with the dispersion among experts' forecasts.

We use scale as a control variable in testing the above hypothesis. Note that the standard deviation of forecast error is commonly known to increase with scale. For example, Conrad (1976) finds that Poisson distribution can be used to represent demand generating processes. Van Ryzin and Mahajan (1999) use a more general distributional assumption, where the standard deviation of demand is a power function of scale, consistent with our model described in §3. Similar to standard deviation of forecast error the dispersion between experts' forecasts could also be increasing in scale. Thus it is possible to observe a positive correlation between the

standard deviation of forecast error and the dispersion among experts' forecasts error just because both variables are correlated with scale (omitted variable bias). Using scale as a control variable in the model avoids this bias.

We also control for the presence of bias in forecasts. Biased forecasts produce non-zero forecast errors (asymptotically) that affect the computation of standard deviation of forecast error (i.e., it is no longer valid to use the absolute value of forecast error as a single point estimate of the standard deviation of forecast error).

3. Model

In this section we develop an empirical model to test the proposed hypothesis. We express the model in general terms because it could be applied to forecasts not only for demand or sales (as we do in this paper), but also for other random variables for which experts' assessments are used.

Let i index the set of random variables. Each random variable is forecasted by several experts and then its realization is observed. Thus, the i -th observation includes the experts' forecasts for and realization of the i -th random variable. For example, each observation might refer to the demand realization for a different product. We do not require the set of experts to be the same for all random variables.

We use the following notation:

X_{ij} : Forecast for random variable i by expert j ;

X_i : Mean of $\{X_{ij}\}$ across experts;

Z_i : Dispersion among experts' forecasts, X_{ij} ;

$$Z_i = \sqrt{\sum_j (X_{ij} - X_i)^2 / (n_i - 1)}, \text{ where } n_i \text{ is the number of experts used for random variable } i;$$

Y_i : Actual realization of random variable i ;

σ_i : Standard deviation of random variable i ;

S_i : Scale factor. We use X_i as the scale factor.

We use a simple regression model to predict random variable Y_i . We assume that all experts are assigned equal weights in computing the forecast. Hence, the only explanatory variable used to predict Y_i is the mean forecast, X_i . However, the mean forecast might be biased. We introduce intercept and slope parameters to account for any bias that might be present. The general form of the forecasting model is

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (1)$$

where ε_i are independent normally distributed error terms with mean 0 and variance σ_i^2 . Here, we use the subscript i for σ_i^2 to model our hypothesis that the variance of the error term differs across observations. Particularly, Hypothesis 1 implies that σ_i is positively correlated with Z_i .

It is possible to test our hypothesis using both non-parametric and parametric tests. For a non-parametric test note that a positive correlation between σ_i and Z_i implies heteroscedastic errors. Hence, a test of heteroscedasticity such as the Goldfeld-Quandt test could be used to test the hypothesis. The Spearman's rank correlation test could also be adapted to test the hypothesis non-parametrically. These tests, explained in §3.1, are advantageous because they do not make any assumptions about the functional form of the relationship between σ_i and Z_i . Hence, conclusions drawn from these tests are robust to erroneous assumptions that might otherwise be made in parametric estimation. The weakness of these tests is that they do not yield a method for estimating σ_i from Z_i .

On the other hand, parametric tests based on regression such as feasible generalized least squares (FGLS) estimation or maximum likelihood estimation (MLE) yield a method for estimating σ_i from Z_i by assuming some functional form of the relationship between σ_i and Z_i . We use a multiplicative functional form as specified in §3.2.

3.1 Non-parametric tests

We perform the Goldfeld-Quandt test for heteroscedasticity (Greene 2003). In order to execute this test we rank the observations based on dispersion among experts' forecasts. The data are then

divided into two equal groups of high and low dispersion and the hypothesis is tested by comparing the estimates of standard error obtained from the regressions on these two groups.

We also perform the Spearman's rank correlation test to estimate the correlation between forecast error and dispersion among experts' forecasts. In this test we first conduct an OLS regression of (1) using all observations. We define the absolute values of the residuals as forecast errors. We rank the forecast error for each observation and the corresponding dispersion among experts' forecasts and compute the correlation between the rankings. Under the null hypothesis of homoscedastic errors this correlation should be zero. On the other hand, under the alternative hypothesis residuals become stochastically more variable as the dispersion among experts' forecasts increases. Hence, the forecast error will be stochastically increasing in the dispersion among experts. Therefore, we expect to observe a positive correlation between forecast error and dispersion among experts' forecasts under the alternative hypothesis.

3.2 Parametric tests

We consider the following regression model with multiplicative heteroscedasticity:

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad (2)$$

$$\varepsilon_i \sim N[0, \sigma_i^2], \quad (3)$$

$$\sigma_i = \lambda (Z_i)^{\gamma_1} (S_i)^{\gamma_2}, \quad (4)$$

where α , β , λ , γ_1 and γ_2 are unknown parameters. Here, ε_i denotes the forecast error and we model the variance of the forecast error as a function of dispersion and scale raised to certain powers. We use this generalized regression model to test the hypothesis that the standard deviation of Y_i is proportional to the dispersion among experts.

This heteroscedasticity model has been well-studied in the econometrics literature. Harvey (1976) uses the term "multiplicative heteroscedasticity" to describe this model and analyzes the statistical properties of various estimators. Besides this model, various other functional forms such as additive, power function, and groupwise have been used in the

econometrics literature to represent heteroscedasticity. Judge et al. (1985) and Greene (2003) provide useful references and detailed descriptions of these models. The heteroscedastic regression model was originally applied to the estimation of household expenditure functions (Battese and Bonyhady 1981).

The multiplicative form (4) is well suited to represent uncertainty in our case for several reasons. First, in application to demand forecasting, it is common to model the standard deviation of demand as having a multiplicative relationship with mean demand. We extend this model by incorporating dispersion as an additional variable. Second, a multiplicative model possesses constant elasticity with respect to the explanatory variables, which is intuitively appealing. Finally, the multiplicative model is more attractive than an additive model from an estimation point of view because (i) the likelihood function is bounded and no problems arise due to estimated variances being negative or zero, and (ii) the error terms in two-step estimation are asymptotically homoscedastic so that the estimated covariance matrix of the vector of parameters, $(\log \lambda, \gamma_1, \gamma_2)$, is consistent. A multiplicative model has the drawback that observations having zero dispersion cannot be included in the estimation when the multiplicative model is transformed using logarithms. We note that it is not only unlikely but also undesirable to have zero dispersion between experts. For example, Hogarth (1978) suggests that the experts selected for a forecasting task should not be strongly correlated with each other.

We estimate model (2)-(4) by maximum likelihood estimation (MLE). Assuming that ε_i are normally distributed, the log-likelihood function is written as:

$$\ln L = -\left(\frac{n}{2}\right) \ln(2\pi) - \frac{1}{2} \sum_i \left[\ln(\sigma_i^2) + \frac{e_i^2}{\sigma_i^2} \right]$$

where e_i are the residuals from (2) and n is the total number of observations. Substituting for variance from (4), we get:

$$\ln L = -\left(\frac{n}{2}\right) \left[\ln(2\pi) + 2\ln\lambda \right] - (\gamma_1) \sum_i \ln Z_i - (\gamma_2) \sum_i \ln S_i - \frac{1}{2\lambda^2} \sum_i \frac{e_i^2}{Z_i^{2\gamma_1} S_i^{2\gamma_2}} \quad (5)$$

We follow the method enumerated in Greene (2003) to estimate the parameters using MLE. First, we fix γ_1 and γ_2 and estimate α , β and λ using the weighted least squares method. Then the likelihood function can be maximized by searching over values of γ_1 and γ_2 . The standard errors of the estimates of the parameters are obtained by computing the inverse of the Hessian matrix.

We also consider an alternative estimation method, feasible generalized least squares (FGLS), for its intuitive appeal. Note that (4) can be written as a log-log regression model as follows:

$$\log (\varepsilon_i)^2 = 2\log \lambda + 2\gamma_1 \log (Z_i) + 2\gamma_2 \log (S_i) + v_i$$

where $v_i = \log(\varepsilon_i^2 / \sigma_i^2)$. Since the true values of α and β are unknown, ε_i is also unknown. Therefore, the estimators of the parameters of (2) and (4) can be obtained by a two-step procedure as follows. First, obtain the ordinary least squares (OLS) residuals for (2), $e_i = y_i - a - bx_i$, where a and b are the OLS estimators for α and β , respectively. Then set up the regression equation

$$\log (e_i^2) \approx 2\log \lambda + 2\gamma_1 \log (Z_i) + 2\gamma_2 \log (S_i) + v_i, \quad (6)$$

where $v_i \approx \log(e_i^2 / \sigma_i^2)$. From (6), the OLS estimators of $\log \lambda$, γ_1 and γ_2 are obtained. These estimators can be used to test the statistical significance of γ_1 and γ_2 ; $\gamma_1 > 0$ would provide evidence for our hypothesis. Further, these estimators can be used to construct revised estimators of α and β , and test their statistical significance.

While we use the FGLS procedure as an alternative, the results reported in this paper are based on MLE. A problem with the FGLS procedure is that if ε_i are normally distributed then v_i are distributed as the logarithm of a χ^2 random variable with one degree of freedom. Hence, v_i have non-zero means.¹ In fact, as the number of observations increases, the asymptotic mean and variance of v_i are $E[v_i] = -1.2704$ and $\text{Var}[v_i] = 4.9348$, respectively. Thus, MLE yields more efficient estimators than the FGLS procedure (Harvey 1976).

¹ We also note that because the computation of (6) is based on e_i rather than ε_i (since the true values of α and β are unknown), v_i are heteroscedastic and autocorrelated. However, as sample size increases v_i will be independent and homoscedastic in the limit.

4. Data Description

We use item-level demand data, firm-level sales data for retail firms, and firm-level sales data for manufacturing firms to test our hypothesis. The first dataset contains style-color level forecasts from individual experts and final demand information for 248 items. The experts were members of a committee specifically constituted to forecast demand. The committee consisted of seven members: the president, a vice president, two designers, and the managers of marketing, production, and customer service (Fisher and Raman 1996). We call this dataset the *Sport Obermeyer dataset*. Table 1 presents summary statistics for the key variables relevant to our study. The average dispersion among experts is 37.6% of forecasts and average forecast error is 60.6 % of forecasts, both values being of the same order of magnitude.

The second and third datasets, which we call the *retail sales dataset* and the *manufacturing sales dataset*, respectively are both taken from the Institutional Brokers Estimate System (I/B/E/S), a Thomson Financial Product, which provides analysts' estimates of various financial measures of U.S. and international companies. The financial measures include sales (SAL), Earnings Per Share (EPS), Book Value Per Share (BPS), Cash Flow Per Share (CPS), and so forth for each fiscal year. Typically, analysts forecast these financial measures for up to three years ahead. I/B/E/S contains two datasets, a Detail History dataset and a Summary History dataset. The Summary History Dataset contains monthly summary statistics such as the mean and standard deviation of the analysts' estimates available up to each month for the abovementioned financial measures. Analysts are expected to update I/B/E/S with any revisions to their forecasts and those who do not issue revised forecasts are presumed by I/B/E/S to stick to their previous forecasts. We merge this dataset with Detail History dataset which contains actual realizations of the financial measures.

We use forecasts of annual sales of firms for the period 1997-2004 for our analysis (Data prior to 1997 contain forecasts of earnings but not of sales). We consider forecasts that are made 90 days or more before the earnings report date (when sales are revealed) for each firm-year

combination. We classify these data based on the time difference between the forecast date and the earnings report date into seven buckets ranging from 3 months prior to earnings report date to 9 or more months prior to earnings report date. Thus, each observation in the dataset is a firm-month-year combination.

I/B/E/S uses a proprietary scheme to classify firms based on business lines into sector/industry/group (SIG) codes. We identified SIG codes that contain firms in the retail and manufacturing sectors as shown in Tables 2a and 2b, respectively. There are 13 SIG codes mapped to the retail sales dataset and 140 SIG codes mapped to the manufacturing sales dataset.

Table 3a summarizes the final retail sales dataset. It consists of 605 observations in each time bucket across 217 firms for the years 1998-2003. Not all firms have data available for each year. However, the same set of firms is represented in each time bucket. In other words, we consider only those firm-year combinations where we have forecasts available up to 9 months prior to the earnings report date. If a firm had forecasts going back up to less than 9 months from the earnings report date for a particular year, then that firm-year was dropped from the dataset to ensure comparability of results across time buckets. Two observations each had to be deleted from the 3-month dataset and the 9-month dataset because there was zero dispersion among experts in these observations. The number of experts in each observation range between 2 and 27, with an average of 5.1. The remaining rows of the table provide summary statistics for all the variables. The average dispersion among experts is 2.88% of forecasts and average forecast error is 4.80% of forecasts.

Table 3b summarizes the manufacturing sales dataset. Here, we have 1171 observations in each time bucket across 479 firms for the same set of years. As with the retail sales dataset, not all firms have data available in each year. Further, only those firms were considered that had forecasts available up to 9 months or more prior to the earnings report date. In each time bucket, a few observations had to be deleted because of zero dispersion between experts. The average

number of experts in an observation is 3.75. The average dispersion among experts is 3.50% of forecasts and average forecast error is 5.84% of forecasts.

5. Results

This section presents our estimation results. We first note that the Goldfeld-Quandt test strongly rejects the null hypothesis of homoscedastic errors ($p < 0.001$) for all three datasets. The Spearman's rank correlation between dispersion among experts and the absolute value of forecast error is 0.385 for the Sport Obermeyer dataset, 0.504 for the retail sales dataset and 0.219 for the manufacturing sales dataset. These correlations are also statistically significant ($p < 0.0001$). Hence, both non-parametric tests support our hypothesis that the standard deviation of forecast error increases with dispersion among experts' forecasts. The rest of this section discusses the results of the parametric tests.

Using MLE, the values of the log likelihood function for the Sport Obermeyer dataset for the full model, the restricted model with $\gamma_1 = 0$, and the restricted model with $\gamma_1 = \gamma_2 = 0$ (i.e., homoscedastic errors) are -1708.0, -1710.4, and -1748.0, respectively. The likelihood ratio test using these values is statistically significant against the restriction $\gamma_1 = 0$ ($p = 0.03$) as well as against the null hypothesis of homoscedastic errors ($p < 0.0001$). Table 4 shows the values of the log likelihood function for the retail sales and the manufacturing sales datasets. We find that the likelihood ratio test is statistically significant for each month: in each case, the χ^2 statistic from this test yields $p < 0.0001$ both against the restriction $\gamma_1 = 0$, and against the null hypothesis of homoscedastic errors.

Tables 5, 6 and 7 present the coefficients' estimates and corresponding standard errors for the three dataset. For the Sport Obermeyer dataset, the estimated coefficient for dispersion among experts, γ_1 , is 0.258 (statistically significant at $p = 0.03$) and the estimated coefficient for scale, γ_2 , is 0.602 (statistically significant at $p < 0.0001$). For the remaining two datasets, all coefficients are statistically significant at $p < 0.0001$. The estimate of γ_1 ranges between 0.143 and 0.360 for the

retail sales dataset, and between 0.159 and 0.193 for the manufacturing sales dataset. The estimate of γ_2 ranges between 0.549 and 0.729, and between 0.790 and 0.837 for the retail sales and the manufacturing sales datasets, respectively. These results consistently support our hypothesis that the standard deviation of Y_i is positively correlated with dispersion among experts' forecasts. Further, dispersion among experts' forecasts has significant explanatory power over and above scale. We also note that the values of coefficients' estimates for our datasets clarify the assumptions in previous research with respect to demand uncertainty. For example, Fisher and Raman (1996) and MacCormack and Verganti (2003) have assumed direct proportionality between the standard deviation of forecast error and dispersion among experts, i.e., $\gamma_1 = 1, \gamma_2 = 0$. Other researchers, e.g., van Ryzin and Mahajan (1999), have assumed that the standard deviation of forecast error is a power function of scale, i.e., $\gamma_1 = 0, \gamma_2 \in [0,1)$. We find that the average values of the estimates for γ_1 and γ_2 across our datasets are 0.201 and 0.715, respectively.

5.1 Longitudinal Analysis of Dataset

We now compare the results of the longitudinal analysis of our dataset using forecasts made 3-9 months before report date. Our observations lead to a conjecture that the estimate of γ_1 increases as we approach the earnings report date, i.e., the closer the forecast date to the report date, the greater is the sensitivity of standard deviation of forecast error to dispersion among experts. However, we offer only one set of observations supporting this conjecture, and it needs to be tested with other datasets in future research.

First, notice from Tables 3a and 3b that the average volatility of sales, measured as absolute percentage error, decreases with the time difference between the forecast date and the earnings report date. It decreases from 6.8% to 2.92% in the retail sales dataset, and from 7.58% to 4.06% in the manufacturing sales dataset. This decrease is intuitive because, with the approach of the earnings report date, a number of uncertainties related to the economy, industry, firm, etc.

that affect sales get resolved or diminish. Correspondingly, the coefficient of variation of dispersion also decreases with time, going from 3.4% to 2.3% for the retail sales dataset, and from 3.9% to 3.1% for the manufacturing sales datasets. This decrease shows that experts agree with each other more and more closely as we approach the earnings report date. Also, consistent with the decrease in the coefficient of variation of sales, we find that the estimate of λ decreases significantly with time. It declines from 0.774 (nine months) to 0.345 (three months) for the retail sales dataset, and from 0.280 (nine months) to 0.137 (three months) for the manufacturing sales dataset. Finally the coefficient of variation of sales measured from our model (i.e., the average of $\lambda(Z_i)^{\gamma_1} (S_i)^{\gamma_2} / (\alpha + \beta X_i)$) decreases with time from 10.2% to 4.7% for the retail sales dataset, and from 11.4% to 7.5% for the manufacturing sales dataset.

The estimates of γ_1 for the two datasets vary slightly with the time difference between the forecast date and the earnings report date. For the retail sales dataset, the estimate increases from 0.166 to 0.282 as we approach the earnings report date. A Bonferroni test on the coefficients' estimates shows that the estimates of γ_1 are statistically similar for months 6 - 9. They are also statistically similar for months 3 - 5. However, the estimate obtained for month 5 is found to be statistically different ($p < 0.0001$) from that obtained for month 6. For the manufacturing sales dataset, the estimates of γ_1 increase from 0.169 to 0.178 as we approach the earnings report date. This increase is consistent with the retail sales dataset but not statistically significant. We believe that these differences may occur due to the release of quarterly earnings reports every 3 months. Quarterly reports provide information about how a firm performed in the previous quarter. This enables experts to adjust their forecasts of annual sales for the firm. It is possible that, as we approach the earnings report date, the dispersion among experts reflects firm-specific uncertainties more accurately so that γ_1 increases in magnitude, as seen in the retail sales dataset. However further research with new datasets and new business scenarios would be needed to analyze the effect of information availability on the coefficients in our model.

6. Conclusions

Standard deviation has been used extensively in models of decision-making, but estimating standard deviation remains a challenge. Recent literature has provided many methods. This paper augments this literature by providing a method to estimate standard deviation using dispersion among point forecasts obtained from multiple experts and managers. It also shows empirically that dispersion among experts' forecasts is a good measure of the volatility of demand. We consider the effect of timing of expert forecasts on the standard deviation of forecast error. Forecasts of sales from I/B/E/S were made about three to nine months before earnings report date. As expected, the standard deviation of forecast error obtained from our model decreases with time as we approach the earnings report date. Tests of hypothesis of the effects of dispersion and scale on standard deviation of forecast error show consistent results in the longitudinal analysis.

Besides the empirical evidence reported in this paper we have tested our hypothesis using a dataset containing forecasts of firms' earnings-per-share (EPS) obtained from the I/B/E/S database. Using about 25,000 observations across 18 years from 1976 to 1993, we found that the standard deviation of forecast error was positively correlated with dispersion among experts in each year. However, we do not include the results of this analysis in the paper because it is well known that EPS data can be managed by firms to reflect analysts' forecasts making it harder to interpret the results in this context. To our knowledge, no such drawbacks have been reported for sales data.

Whereas we considered only two explanatory variables, dispersion and scale, it is possible to extend our model to a larger set of variables. Uncertainty of demand might be correlated with other factors such as firm-specific risk, market uncertainty, macroeconomic conditions, and product characteristics (e.g., basic versus fashion apparel). For example, given the mean forecast for an item and dispersion among experts, the standard deviation of the forecast error may increase with the volatility of a market index. To a certain extent dispersion among experts would capture uncertainties related to such factors since experts *should consider all*

available relevant information (Makridakis and Wheelwright 1987) in forecasting demand. However, our model can easily be augmented by including additional factors in the multiplicative model (4) to estimate their effects on demand uncertainty.

One limitation of our study is that we assume all analysts to be equally capable. Thus, we do not consider the effects of the composition of analysts' committees to derive our results. Given suitable data, it would be easy to extend our methodology to analyze the quality of predictions made by different analysts and, thus, compare their capabilities. This would require forecasts of several random variables by the same set of analysts. This research can also be extended by applying our model to different datasets in future to confirm the findings in this paper and analyze dispersion among experts' forecasts.

References

- Ajinkya, B. J., M. J. Gift. 1985. Dispersion of Financial Analysts' Earnings Forecasts and the (Option Model) Implied Standard Deviations of Stock Returns. *The Journal of Finance*. **40**(5) 1353-1365.
- Armstrong, S.J. 1986. Research on Forecasting: A Quarter-Century Review, 1960-1984. *Interfaces*. **16**(1) 89-103.
- Battese, G. E., B. P. Bonyhady. 1981. Estimation of Household Expenditure Functions: An Application of a Class of Heteroscedastic Regression Models. *The Economic Record*. **57** 80-85.
- Conrad, S. A. 1976. Sales Data and Estimation of Demand. *Operations Research Quarterly*. **27** 123-127.
- Dalkey, N., O. Helmer. 1963. An Experimental Application of the Delphi Method to the Use of Experts. *Management Science*. **9**(3) 458-467.
- Fisher, M. L., A. Raman. 1996. Reducing the Cost of Demand Uncertainty through Accurate Response to Early Sales. *Operations Research*. **44**(1) 87-99.
- Flugstad, A. R., P. D. Windschitl. 2003. The Influence of Reasons on Interpretations of Probability Forecasts. *Journal of Behavioral Decision Making*. **16**(2) 107-126.
- Givoly, D., J. Lakonishok. 1984. Properties of Analysts' Forecasts of Earnings: A Review and Analysis of the Research. *Journal of Accounting Literature*. **3** 117-152.
- Greene, W. H. 2003. *Econometric Analysis*. Fifth Edition. Prentice Hall, New Jersey.
- Harvey, A. C. 1976. Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrica*. **44**(3) 461-465.
- Hogarth, R.M. 1978. A Note on Aggregating Opinions. *Organizational Behavior and Human Performance*. **21**(1) 40-52.
- Judge, G. G., W. E. Griffiths, R. Carter Hill, H. Lutkepohl, T. C. Lee. 1985. *The Theory and Practice of Econometrics*. Second Edition. New York: John Wiley & Sons.

- MacCormack, A., R. Verganti. 2003. Managing the Sources of Uncertainty: Matching Process and Context in Software Development. *The Journal of Product Innovation and Management*. **20** 217-232.
- Makridakis, S., S. C. Wheelwright. 1987. *The Handbook of Forecasting: A Manager's Guide*, Second Edition. New York: John Wiley & Sons.
- Malkiel, B. G. 1982. Risk and Return: A New Look. NBER Working Paper No. W0700.
Weblink: <http://ssrn.com/abstract=227532>.
- Milkovich G. T., A. J. Annoni, T. A. Mahoney. 1972. The Use of the Delphi Procedures in Manpower Forecasting, *Management Science*, **19**(4) 381-388.
- McKenney J.L., T.H.Clark. 1994. Campbell Soup Company: A Leader in Continuous Replenishment Innovations, Harvard Business School, Case # 9-195-124, Cambridge, MA.
- Munier F., P. Ronde. 2001. The Role of Knowledge Codification in the Emergence of Consensus under Uncertainty: Empirical Analysis and Policy Implications, *Research Policy*. **30**(9) 1537-1551.
- Soll, J. B., J. Klayman. 2004. Overconfidence in Interval Estimates. *Journal of Experimental Psychology*. **30** 299-314.
- Summers, B., T. Williamson, D. Read. 2004. Does method of Acquisition affect the Quality of Expert Judgment? A Comparison of education with on-the-job learning. *Journal of Occupational & Organizational Psychology*. **77**(2) 237-258.
- van Ryzin, G., S. Mahajan. 1999. On the relationship between inventory costs and variety benefits in retail assortments. *Management Science*. **45**(11) 1496-1509.

Table 1: Summary Statistics for Demand Data from Sport Obermeyer

(Number of items = 248, Number of experts who forecast for each item = 7)

Variable	Mean	Standard Deviation	Minimum	Maximum
Actual Sales, Y_i	382.15	348.29	4.00	1720.00
Average Forecast (Scale), X_i	367.45	214.71	60.00	1151.67
Coefficient of variation of forecasts, Z_i/X_i (%)	37.6	14.4	5.6	95.0
Absolute Percentage Error, $ Y_i - X_i / X_i$ (%)	60.6	50.9	0.4	380.0

Table 2a: Mapping of SIG codes to Retail Dataset

SIG Codes	Sector Name	Industry Name	Examples of Group Names
40301, 40302	Consumer Services	Retailing – Foods	Grocery Chains, Food Wholesalers.
40401 – 40406, 40408 – 40413	Consumer Services	Retailing - Goods	Department Stores, Specialty Retailers, Variety Chains, Drug Chains, Electronics Chains, etc.

Table 2b: Sample Mapping of SIG codes to Manufacturing Dataset

SIG Codes	Sector Name	Examples of Industry and Group Names
30001 – 39901	Consumer Non-Durables	Apparel, Shoes, Cosmetics, Packaged foods, Beverages, Toys and games, etc.
50101 – 59901	Consumer Durables	Auto, Auto Parts, Home furnishings, Tools and hardware, etc.
80101 – 80302, 80701 – 81102	Technology	Computer manufacturers, Semiconductors, Electronic devices, Photo-optical equipment, etc.
90201 – 90501, 90701 – 91201	Basic Industries	Chemical, Containers, Metal Fabrication, Fertilizers, Steel, Textile, etc.
100101 – 109901	Capital Goods	Defence, Auto OEMs, Electrical, Machinery, Building material, etc.

Table 3a: Summary Statistics for Retail Sales Data

Variable	Statistic	Time difference between forecast date and earnings report date (in months)						
		3	4	5	6	7	8	9
Number of firms represented		215	217	217	217	217	217	215
Number of observations		603	605	605	605	605	605	603
Number of experts per observation	Mean	5.3	5.2	5.2	5.2	5.0	4.9	4.9
	Std. Deviation	3.8	3.8	3.8	3.7	3.5	3.5	3.3
	Minimum	2	2	2	2	2	2	2
	Maximum	24	24	27	26	25	25	22
Actual Sales, Y_i	Mean	7761	7737	7737	7737	7737	7737	7737
	Std. Deviation	21328	21296	21296	21296	21296	21296	21296
	Minimum	6.9	6.9	6.9	6.9	6.9	6.9	6.9
	Maximum	258681	258681	258681	258681	258681	258681	258681
Actual Forecast (Scale), X_i	Mean	7761	7754	7757	7754	7757	7755	7802
	Std. Deviation	21330	21308	21339	21297	21306	21339	21550
	Minimum	8.8	9.1	9.1	8.5	7.9	7.9	9.5
	Maximum	258201	257821	257609	257043	257782	259520	267295
Coefficient of variation of forecasts, Z_i/X_i (%)	Mean	2.3	2.5	2.7	2.8	3.2	3.3	3.4
	Std. Deviation	3.7	3.2	3.6	3.7	4.6	5.6	5.6
	Minimum	0.02	0.04	0.06	0.02	0.0023	0.0013	0.0023
	Maximum	41.1	28.5	33.0	34.0	46.0	80.0	80.0
Absolute Percentage Error, $ Y_i - X_i / X_i$ (%)	Mean	2.92	3.53	4.08	4.70	5.70	6.11	6.80
	Std. Deviation	4.52	4.90	5.67	6.73	8.27	8.80	9.69
	Minimum	0.000	0.001	0.007	0.001	0.001	0.011	0.0001
	Maximum	47.53	49.40	49.40	65.78	80.90	83.30	83.30

Table 3b: Summary Statistics for Manufacturing Sales Data

Variable	Statistic	Time difference between forecast date and earnings report date (in months)						
		3	4	5	6	7	8	9
Number of firms represented		479	479	479	478	479	479	479
Number of observations		1165	1165	1165	1167	1164	1165	1171
Number of experts per observation	Mean	4.03	3.94	3.88	3.52	3.76	3.72	3.38
	Std. Deviation	2.54	2.46	2.39	2.20	2.37	2.33	2.08
	Minimum	2	2	2	2	2	2	2
	Maximum	22	22	22	24	22	21	20
Actual Sales, Y_i	Mean	7208	7208	7208	7139	7214	7208	7264
	Std. Deviation	19077	19077	19077	18918	19084	19077	19076
	Minimum	18.3	18.3	18.3	18.3	18.3	18.3	18.3
	Maximum	186763	186763	186763	186763	186763	186763	186763
Actual Forecast (Scale), X_i	Mean	7001	7039	7049	6964	7061	7050	7062
	Std. Deviation	17593	17801	17808	17546	17857	17852	17822
	Minimum	17.40	16.73	16.73	16.73	16.73	16.73	16.73
	Maximum	164174	164101	186763	163999	168862	168058	167798
Coefficient of variation of forecasts, Z_i/X_i (%)	Mean	3.1	3.2	3.3	3.5	3.7	3.8	3.9
	Std. Deviation	4.9	4.6	4.7	4.8	5.0	5.0	5.0
	Minimum	0.0064	0.0049	0.0049	0.0049	0.01	0.01	0.01
	Maximum	52	52	52	52	52	52	52
Absolute Percentage Error, $ Y_i - X_i / X_i$ (%)	Mean	4.06	4.74	5.11	5.71	6.60	7.10	7.58
	Std. Deviation	6.90	7.14	7.35	7.68	8.33	8.78	9.15
	Minimum	0.0001	0.0001	0.0007	0.0080	0.0080	0.0008	0.0007
	Maximum	79	79	79	82	82	82	79

Table 4: Loglikelihood Values of the MLE Results for Retail and Manufacturing Sales Datasets

	Time difference between forecast date and earnings report date (in months)						
	3	4	5	6	7	8	9
Retail Sales Dataset							
Full model	-3574	-3666	-3715	-3882	-3987	-4016	-4050
$\gamma_1 = 0$	-3657	-3735	-3810	-3905	-4011	-4047	-4082
$\gamma_1 = 0$ and $\gamma_2 = 0$	-4564	-4621	-4655	-4752	-4753	-4767	-4902
Manufacturing Sales Dataset							
Full model	-7378	-7471	-7538	-7602	-7738	-7813	-7874
$\gamma_1 = 0$	-7479	-7557	-7611	-7670	-7803	-7875	-7934
$\gamma_1 = 0$ and $\gamma_2 = 0$	-10620	-10645	-10715	-10790	-10823	-10851	-10916

Table 5: MLE Results for the Sport Obermeyer Data

Parameter	Sport Obermeyer Data	
	Estimate	Std. Error
α	-17.18	19.748
β	1.09**	0.081
λ	2.25*	0.974
γ_1	0.258*	0.121
γ_2	0.602**	0.114

*, **: statistically significant at 0.05, 0.01.

Table 6: MLE Results for the Longitudinal Study of Retail Sales Data

Parameter	Time difference between forecast date and earnings report date (in months)						
	3	4	5	6	7	8	9
α	-0.827 (0.526)	-0.525 (0.542)	-0.427 (0.581)	-0.803 (0.995)	0.650 (1.357)	0.886 (1.469)	1.261 (1.763)
β	1.002** (0.002)	1.000** (0.002)	0.999** (0.002)	1.002** (0.003)	1.000** (0.003)	1.000** (0.003)	1.000** (0.003)
λ	0.345** (0.058)	0.350** (0.060)	0.480** (0.080)	0.322** (0.049)	0.560** (0.090)	0.652** (0.105)	0.774** (0.108)
γ_1	0.282** (0.024)	0.290** (0.027)	0.360** (0.028)	0.160** (0.027)	0.143** (0.026)	0.147** (0.025)	0.166** (0.020)
γ_2	0.611** (0.029)	0.615** (0.030)	0.549** (0.030)	0.729** (0.029)	0.685** (0.029)	0.670** (0.029)	0.646** (0.026)
Average of [Dispersion / Mean Forecast] (%)	2.3	2.5	2.6	2.8	3.2	3.3	3.4
Average CV of Sales (%)	4.7	5.4	5.9	7.0	9.3	9.5	10.2

Note: (i) **: statistically significant at 0.01. (ii) Average CV (coefficient of variation) of sales is computed using (4) as the average of $\lambda(Z_i)^{\gamma_1} (S_i)^{\gamma_2} / \alpha + \beta X_i$ across all observations.

Table 7: MLE Results for the Longitudinal Study of Manufacturing Sales Data

Parameter	Time difference between forecast date and earnings report date (in months)						
	3	4	5	6	7	8	9
α	0.085 (0.409)	0.349 (0.454)	0.341 (0.497)	0.182 (0.551)	-0.13 (0.616)	0.021 (0.696)	0.204 (0.692)
β	1.008** (0.002)	1.008** (0.002)	1.007** (0.002)	1.006** (0.003)	1.008** (0.003)	1.008** (0.003)	1.007** (0.003)
λ	0.137** (0.016)	0.181** (0.020)	0.197** (0.020)	0.210** (0.024)	0.240** (0.029)	0.280** (0.030)	0.280** (0.030)
γ_1	0.178** (0.016)	0.193** (0.018)	0.176** (0.018)	0.169** (0.018)	0.167** (0.018)	0.159** (0.017)	0.169** (0.018)
γ_2	0.837** (0.020)	0.800** (0.020)	0.800** (0.020)	0.810** (0.020)	0.790** (0.020)	0.790** (0.020)	0.790** (0.020)
Average of [Dispersion / Mean Forecast] (%)	3.1	3.2	3.3	3.5	3.8	3.8	3.9
Average CV of Sales (%)	7.5	8.0	8.6	9.1	10.2	11.0	11.4

Note: (i) **: statistically significant at 0.01. (ii) Average CV (coefficient of variation) of sales is computed using (4) as the average of $\lambda(Z_i)^{\gamma_1} (S_i)^{\gamma_2} / \alpha + \beta X_i$ across all observations.