Electronic Companion to:

A Bayesian Model for Sales Forecasting at Sun Microsystems

Phillip M. Yelland, Shinji Kim, Renée Stratulate

References to figures and tables specific to this electronic companion are distinguished by the prefix "EC-"; other references refer to the printed paper.

Related Work

There is a wealth of literature documenting the biases and errors associated with judgment-based forecasting—c.f. (McGlothlin 1956, Tversky and Kahneman 1974, Wright and Ayton 1986, Bolger and Harvey 1998), for example. Mentzer and Bienstock (1998) and Tyebjee (1987) point out that in addition to these problems, judgmental sales forecasts could be distorted by other factors, such as organizational pressures. Many researchers and practitioners have sought to combine judgmental and statistical forecasting techniques. Broadly speaking, such efforts may be categorized under the headings: Adjustment, combination, correction, judgmental bootstrapping, and Bayesian methods.

Adjustment

Articles, such as Sanders and Manrodt (1994), Fildes and Goodwin (2007), Franses and Legerstee (2009), and Fildes et al. (2009), provide evidence of the widespread manual adjustment of statistical forecasts *ex post facto*—frequently, by "eyeball" analysis of forecast graphs, as Bunn and Wright (1991) and Webby and O'Connor (1996) describe. This approach has intuitive appeal. However, Armstrong and Collopy (1998) point out that if we carry out adjustments in an unstructured or undisciplined fashion, we risk simply reintroducing the distortions of judgmental forecasting. In fact, the accounts of Fildes and Goodwin (2007), Franses and Legerstee (2009), and Fildes et al. (2009) suggest that more often than not in practice, adjustments *reduce* the accuracy of a statistical forecast. Bunn and Wright (1991) also aver that without sufficient ancillary detail, seeing how an apparently arbitrary adjustment was justified can be difficult, and can lead to possible contention within the organization employing the forecast.

Combination

An alternative approach, as Blattberg and Hoch (1990), Webby and O'Connor (1996), Lawrence et al. (2006), and Franses (2008) discuss, is to originate statistical and judgmental forecasts independently and use a mechanical procedure to combine them. This builds upon work on forecast combination, which began with Bates and Granger (1969), and which Clemen (1989) and Timmermann (2006) survey. Combination is generally effected by taking a weighted sum of the values of the constituent forecasts in each period. Many methods have been proposed to estimate the combination weights: Granger and Ramanathan (1984) suggest least-squares regression; Bates and Granger (1969) compute weights based on the forecast errors of the constituent forecasts, and Diebold and Pauly (1990) use Bayesian shrinkage regression. In practice, a simple unweighted average of the component forecasts, as Blattberg and Hoch (1990) and Armstrong (2001a) note, performs consistently well.

Timmermann (2006) lists several factors that recommend forecast combination: It synthesizes the information sets used to produce the component forecasts, it dilutes bias in individual forecasts, and it increases robustness with respect to model misspecification and structural breaks. Set against this, Timmermann also notes that estimated combination weights can be very unstable in practice, which helps explain the remarkably good relative performance of the simple average. In principle, concludes Diebold (1989), where the information sets underlying the component forecasts are available, it is always preferable to construct a single encompassing forecast model, rather than simply to combine the forecasts.

Correction

Rather than combining a statistical and a judgmental forecast, some authors, including Theil (1971), Ahlburg (1984), Moriarty (1985), Elgers et al. (1995), and Goodwin (1996, 2000), have explored statistical methods for correcting judgmental forecasts in the light of observed outcomes. Generally, such methods are based on Theil's (1971) "optimal linear correction," which involves regressing observed outcomes on forecasts, using the estimated regression coefficients to produce a revised prediction from new forecasts; Goodwin (1997) accommodates time-varying coefficients using a weighted regression.

A technique related to statistical correction originates with Lindley (1983), and is applied explicitly to time-series forecasting by West and Harrison (1997, sec. 16.3.2). Lindley's methodology is an example the *supra-Bayesian* approach to the reconciliation of expert opinion developed by Pankoff and Roberts (1968) and Morris (1974, 1983). Here, the value of a judgmental forecast is construed as a linear function of the actual value, and Bayesian updating is used to produce a revised forecast.

Judgmental Bootstrapping

Researchers, particularly in psychology, have long sought to capture judgmental reasoning in a tractable mathematical form. Efforts centering on linear models date back at least to Hughes (1917); Meehl (1957), Hammond et al. (1964), Hursch et al. (1964), and Tucker (1964) also made important contributions. Surveying this work, Dawes (1971) coined the term *bootstrapping* to describe the process by which an expert's judgment is modeled by a linear expression involving the environmental factors (usually referred to as *cues*) that the expert considers. Recent authors of works that explore the application of such a process to judgment-based forecasting, including Armstrong (2001b), O'Connor et al. (2005), and Batchelor and Kwan (2007), use the qualified term *judgmental* bootstrapping to avoid confusion with the (quite distinct) statistical bootstrap technique that Efron (1979) developed in the late 1970s. Armstrong (2001b) also applies the term to models that go beyond simple linear combination of cues. We use the qualified form "judgmental bootstrapping" in this paper—more distinctive alternative terms, such as "the statistical approach," "actuarial modeling" (both of which appear in the clinical literature) or the (mildly grandiloquent) "paramorphic representation" of Hoffman (1960) would doubtless invite even greater confusion.

Evidence for the efficacy of judgmental bootstrapping in forecasting is mixed: Ashton et al. (1994) find a bootstrap model outperformed by a statistical forecasting model, and Åstebro and Elhedhli (2006) and Batchelor and Kwan (2007) cannot conclude that a bootstrap model forecasts more accurately than the experts it represents. Lawrence and O'Connor (1996) and Fildes et al. (2009) assert that bootstrapping is less effective in the context of time-series extrapolation, where cue information tends to be autocorrelated; the "error bootstrapping" technique developed in response by Fildes (1991) seeks to model the *errors* in a judgmental forecast, much like the forecast correction approach described above.

Bayesian Methods

The Bayesian paradigm of statistical inference (Gelman et al. 2003), with its incorporation of subjective information in the form of prior distributions, seems a natural means of combining judgmental and statistical elements in forecasting. Indeed, a substantial number of Bayesian models have been devised for product demand, both at the inventory level (Silver 1965, Hill 1997, Dolgui and Pashkevich 2008a,b) and in the aggregate (Lenk and Rao 1990, Montgomery 1997, Moe and Fader 2002, Neelamegham and Chintagunta 1999, van Heerde et al. 2004, Neelamegham and Chintagunta 2004, Lee et al. 2008)—among many others. Sales applications also feature prominently in the seminal work of Pole et al. (1994) and West and Harrison (1997) on Bayesian forecasting.

Bunn and Wright (1991) note that despite the apparent attractions of Bayesian modeling, and research literature that dates back several decades, few Bayesian models with judgmental priors are used routinely in forecast applications. They suggest that the chief impediment is the expense (in time and effort) of repeatedly eliciting subjective priors of sufficient quality—a point reinforced by researchers, such as Wright and Ayton (1986), who highlight the difficulties involved in obtaining reliable judgmental priors. Many of the models cited in the previous paragraph circumvent the need for informative priors by relying on a *hierarchical* (Gelman and Hill 2006) structure to pool information from analogous historical situations to produce forecasts.

Forecast Performance

To demonstrate the efficacy of the SLFS, we compare its performance with that of alternative forecasting methods, using the demand histories of a sample of Sun's products.

Setup

We conduct the test using 32 products, representing a cross section of Sun's recent product lines. Of these products, we randomly select 27 as "training data" to calibrate the forecasting methods, and produce forecasts for the remaining 5 (the latter so-called *holdout* products are excluded from the training data). The relatively large number of products in the training data collection is necessary to ensure reliable operation of forecast methods Sys, DLM, and CJudg, which require a representative set of products from each of Sun's product groups. We prepare point forecasts at horizons of one, two, and three quarters for each quarter of the holdout products' demand histories, yielding 77 forecast values for each method at each horizon. Four forecast methods are compared in the test:

Sys This forecasting method relies on *posterior predictive distributions* of sales calculated by the SLFS, the standard approach to forecasting with Bayesian models (Neelamegham and Chintagunta 1999, 2004). Formally, the posterior distribution of the sales y_{t+h} at horizon h is the conditional distribution $p(y_{t+h}|D_t)$, where D_t denotes the data (i.e., demand histories and judgmental forecasts) available at time t. As Albert (2008) demonstrates, this conditional distribution is readily approximated using a Gibbs sampler such as that specified in the appendix, which implements the Bayesian updating step of Figure 2 by simulating from the joint distribution of the parameters in the model in light of the available data. For the point forecast $\hat{y}_{t+h|t}$ (the forecast for y_{t+h} made in period t), we use the mean of the posterior predictive distribution of y_{t+h}.

DLM This method is a straightforward implementation of a univariate dynamic linear model (W&H, chap. 5–8). Following W&H's prescription, the model comprises a second-order polynomial ("local level with trend") and four-period seasonal component. As Gardner and McKenzie (1989) suggest, to improve forecasts at over longer horizons, the trend in the polynomial component is *damped*, so that it decays over time; the design and evolution matrices necessary to accomplish this are adapted from analogous structures described for the closely related *single source of error* (SSOE) structural time-series models (Hyndman et al. 2008, p. 48).

Formulae for updating and forecasting with this DLM are standard: We use the "unknown, constant variance" results (W&H, p. 111), which incorporate the estimation of a time-invariant observation noise variance in Equation (1a). Multiple *discount factors* are used to specify the evolution noise component in Equation (1b) (c.f. *op. cit.* pp. 196–198 for details). Discount and damping factors are derived using a grid search based on forecast accuracy for a sample of the calibration set, and initial conditions are set from the same corrected priors used for the Sys method, using a procedure similar to that described by Hyndman et al. (2008). All components in the state are uncorrelated in the prior, so that the matrix C_0 in Equation (1c) is diagonal.

Judg This method simply reproduces the company's judgmental forecast, from which the priors for methods Sys and DLM are derived.

CJudg For this method, the company's judgmental forecast is "corrected" using Theil's (1971) optimal linear correction. This involves regressing actual sales on predicted sales for products in the calibration set, using the estimated coefficients to compute revised prediction from forecasts in the holdout set; Theil (1971, p. 34) and Goodwin (2000) provide detailed discussions. Separate regressions are calculated for each forecast horizon and each product category identified for prior correction.

Test Results

Table EC-1 summarizes the performance of the candidate methods in the forecasting test. We have used the *mean absolute scaled error* (MASE) of Hyndman and Koehler (2006) as a performance metric; the appendix defines the MASE, and sets out the considerations that lead to its adoption. MASEs are given for each method at each forecast horizon—a smaller entry indicates better performance. As the table shows, the forecast system consistently exhibits superior overall performance, with the purely statistical DLM method generally turning in the worst performance, and the corrected judgmental forecast also consistently

	Horizon		
	1	2	3
Sys	0.53 (1)	0.84 (1)	0.80 (1)
DLM	0.73 (3)	1.13 (4)	0.97 (4)
Judg	0.73 (4)	1.10 (3)	0.96 (3)
CJudg	0.66 (2)	1.00 (2)	0.94 (2)

Table EC-1: In this table, the performance of each of the forecasting methods in the comparison is summarized by its mean absolute scaled error at each forecast horizon. The rank (smallest to largest) of each method at each horizon is recorded in parentheses after the corresponding entry.

outperforming its uncorrected counterpart. The MASE is defined such that a value less than 1.00 indicates a performance better (on average) than that of the "naïve" random-walk forecast, which simply uses the most recent observation to produce a forecast. Thus, according to the table, only the Sys method consistently improves on the benchmark.

For a more detailed perspective, Figure EC-1 is a box-and-whisker plot (Tukey 1977) of the distribution of absolute scaled errors (ASEs) of the forecast methods at each forecast horizon. Although the distributions of the ASEs are clearly skewed, the plot broadly confirms the superiority of the Sys method established by the MASEs.

Appendix

Measures of Forecast Accuracy

As a way of measuring the quadratic loss criterion that pervades statistical inquiry, the *mean square error* and its related metric, the *root mean square error* (we refer to both metrics jointly as the (R)MSE) have long been a staple of academic research in forecasting (Granger and Newbold 1973). Unfortunately, although the (root) mean square error is analytically attractive, researchers, such as Armstrong and Collopy (1992), point to a number of practical problems with its use: (1) Performance with respect to the (R)MSE may be significantly affected by outliers. Such problems may be ameliorated by eliminating outliers, although this might draw the objectivity of the procedure into question. (2) More seriously, the (R)MSE is inherently scale dependent, in that its magnitude depends not only on forecast accuracy, but also on the level of the underlying series; *ceteris paribus*, a forecast 10 percent in excess of an actual value of 1,000,000 will result in a substantially greater (R)MSE than one 10 percent above an actual of 1,000. This largely invalidates sum-

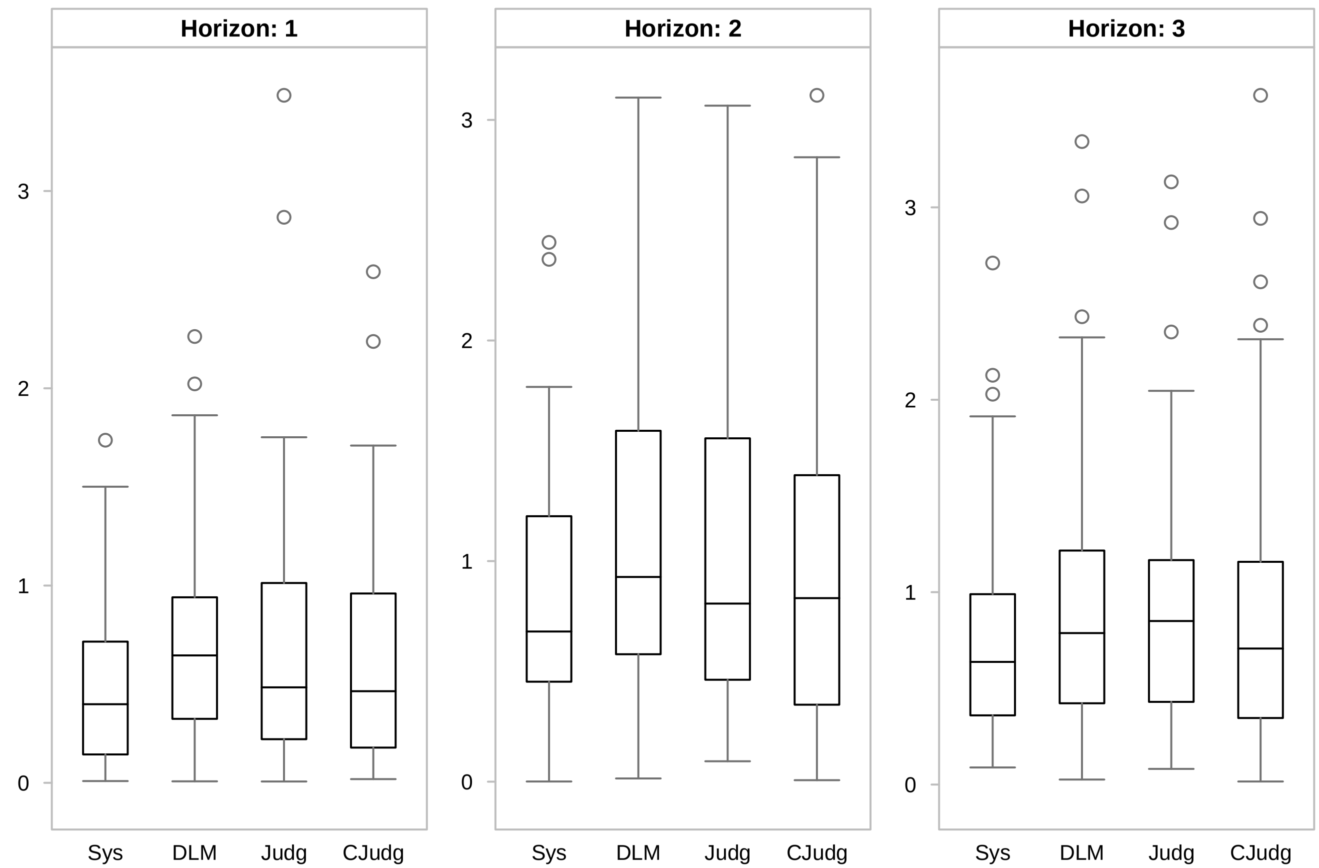

Figure EC-1: The box-and-whisker plots in each panel illustrate the distributions of absolute scaled errors by model and forecast horizon in the performance test.

mary measures based on the (R)MSE of performance across a heterogeneous collection of series; Chatfield (1988), for example, points out that Zellner's (1986) analysis was unduly affected by the outsized contribution to summary MSE metrics of only 5 out of 1,001 test series. Because the series in this test described in the paper vary in maximum between 200 and 4,000, the (R)MSE is unsuited to this application.

The *mean absolute percentage error* (MAPE) favored by practitioners, which expresses error as a fraction of the associated actual value, avoids the scale dependency of the (R)MSE. However, the MAPE also has disadvantages. (1) In contradistinction to the (R)MSE, summary MAPE measures may be skewed by *small* actuals; indeed, the MAPE is infinite for an actual value of 0. Some researchers, such as Coleman and Swanson (2007), have suggested taking logarithms as a way of mitigating the problem, although this makes the resulting metric more difficult to interpret. (2) The MAPE exhibits a counterintuitive asymmetry; a forecast of 5 units on an actual of 10 produces an absolute percentage error of 50 percent, whereas a forecast of 10 units on an actual of 5 gives an APE of 100 percent. Attempts to amend the MAPE to overcome this problem (Makridakis 1993) have had limited success (Koehler 2001).

In light of these problems, some authors, including Fildes (1992) and Armstrong and Collopy (1992), have proposed using metrics based on *relative* absolute errors (RAEs). These are absolute forecast errors divided by the corresponding error from a benchmark method (normally the "naïve" or "random walk" method, which simply repeat the last observation). However, Hyndman and Koehler (2006) point out that RAE-based metrics suffer from some of the same problems as the (R)MSE and MAPE: They are sensitive to outliers, and may be skewed by small benchmark forecast errors (again, the RAE is infinite if the forecast error is 0); remedies involving outlier elimination and log transformation are also subject to the same criticisms.

Hyndman and Koehler (2006) propose the MASE as a robust, scale-independent metric that largely avoids the problems set out above. Formally, for series $y_1, \ldots, y_T$, and denoting by $\hat{y}_{t+h|t}$ the forecast for y_{t+h} made in period t, the absolute scaled error ASE_{th} is defined:

$$\mathrm{ASE}_{th} = \frac{|\hat{y}_{t+h|t} - y_{t+h}|}{\frac{1}{T-h}\sum_{t'=1}^{T-h} |y_{t'+h} - y_{t'}|}$$

Then the mean absolute scaled error, MASE_h at horizon h for the entire series is simply the mean $T^{-1}\sum_{t=1}^{T} \mathrm{ASE}_{th}$, and a summary metric for a collection of series may be calculated by taking the mean of the ASEs across all the series. In recommending the MASE, Hyndman and Koehler (2006) note that normalization with respect to the benchmark forecast error confers scale independence, while use of the in-sample aggregate in the denominator of scaled errors makes the MASE metric more stable than those based on relative errors.

References

Ahlburg, D. A. 1984. Forecast evaluation and improvement using Theil's decomposition. *J. Forecasting* **3** 345–351.

Albert, J. 2008. *Bayesian Computation with R*. Springer, New York.

Armstrong, J., F. Collopy. 1992. Error measures for generalizing about forecasting methods—empirical comparisons. *Internat. J. Forecasting* **8** 69–80.

Armstrong, J. M. 2001a. Combining forecasts. J. M. Armstrong, ed., *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer, Norwell, MA, 417–439.

Armstrong, J. M. 2001b. Judgmental bootstrapping: Inferring experts' rules for forecasting. J. M. Armstrong, ed., *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer, Norwell, MA, 171–192.

Armstrong, J. M., F. Collopy. 1998. Integration of statistical methods and judgment for time series forecasting: Principles from empirical research. G. Wright, P. Goodwin, eds., *Forecasting with Judgment*. John Wiley & Sons, New York, 269–293.

Ashton, A. H., R. H. Ashton, M. N. Davis. 1994. White-collar robotics: Levering managerial decision making. *California Management Rev.* **37** 83–109.

Åstebro, T., S. Elhedhli. 2006. The effectiveness of simple decision heuristics: Forecasting commercial success for early-stage ventures. *Management Sci.* **52** 396–409.

Batchelor, R., T. Y. Kwan. 2007. Judgemental bootstrapping of technical traders in the bond market. *Internat. J. Forecasting* **23** 427–445.

Bates, J. M., C. W. J. Granger. 1969. The combination of forecasts. *Oper. Res. Quart.* **20** 451–468.

Blattberg, R. C., S. J. Hoch. 1990. Database models and managerial intuition: 50% model + 50% manager. *Management Sci.* **36**(8) 887–899.

Bolger, F., N. Harvey. 1998. Heuristics and biases in judgmental forecasting. G. Wright, P. Goodwin, eds., *Forecasting with Judgment*. John Wiley & Sons, New York, 113–137.

Bunn, D., G. Wright. 1991. Interaction of judgemental and statistical forecasting methods: Issues & analysis. *Management Sci.* **37**(5) 501–518.

Chatfield, C. 1988. Apples, oranges and mean square error. *J. Forecasting* **4** 515–518.

REFERENCES

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. *Internat. J. Forecasting* **5** 559–583.

Coleman, C. D., D. A. Swanson. 2007. On MAPE-R as a measure of cross-sectional esimation and forecast accuracy. *J. Econ. Social Measurement* **32** 219–233.

Dawes, R. M. 1971. A case study of graduate admissions: Application of three principles of human decision making. *Amer. Psych.* **26** 180–188.

Diebold, F. X. 1989. Forecast combination and encompassing: Reconciling two divergent literatures. *Internat. J. Forecasting* **5** 589–592.

Diebold, F. X., P. Pauly. 1990. The use of prior information in forecast combination. *Internat. J. Forecasting* **6** 503–508.

Dolgui, A., M. Pashkevich. 2008a. Demand forecasting for multiple slow-moving items with short requests history and unequal demand variance. *Internat. J. Production Econom.* **112** 885–894.

Dolgui, A., M. Pashkevich. 2008b. On the performance of binomial and beta-binomial models of demand forecasting for multiple slow-moving inventory items. *Comput. Oper. Res.* **35** 893–905.

Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

Elgers, P. T., H. L. May, D. Murray. 1995. Note on adjustments to analysts' earning forecasts based upon systematic cross-sectional components of prior-period errors. *Management Sci.* **41** 1392–1396.

Fildes, R. 1991. Efficient use of information in the formation of subjective industry forecasts. *J. Forecasting* **10** 597–617.

Fildes, R. 1992. The evaluation of extrapolative forecasting methods. *Internat. J. Forecasting* **8** 81–98.

Fildes, R., P. Goodwin. 2007. Good and bad judgement: Lessons from four companies. *Foresight: Internat. J. Appl. Forecasting* **8** 5–10.

Fildes, R., P. Goodwin, M. Lawrence, K. Nikolopoulos. 2009. Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *Internat. J. Forecasting* **25** 3–23.

Franses, P. 2008. Merging models and experts. *Internat. J. Forecasting* **24** 31–33.

REFERENCES

Franses, P. H., R. Legerstee. 2009. Properties of expert adjustments on model-based SKU-level forecasts. *Internat. J. Forecasting* **25** 35–47.

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

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

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

Goodwin, P. 1996. Statistical correction of judgmental point forecasts and decisions. *Omega: Internat. J. Management Sci.* **24** 551–559.

Goodwin, P. 1997. Adjusting judgemental extrapolations using Theil's method and discounted weighted regression. *J. Forecasting* **16** 37–46.

Goodwin, P. 2000. Correct or combine? Mechanically integrating judgmental forecasts with statistical methods. *Internat. J. Forecasting* **16** 261–275.

Granger, C. W. J., P. Newbold. 1973. Some comments on the evaluation of economic forecasts. *Appl. Econom.* **5** 35–47.

Granger, C. W. J., R. Ramanathan. 1984. Improved methods of combining forecasts. *J. Forecasting* **3** 197–204.

Hammond, K. R., C. J. Hursch, F. J. Todd. 1964. Analyzing the components of clinical inference. *Psych. Rev.* **71** 255–262.

Hill, R. 1997. Applying Bayesian methodology with a uniform prior to the single period inventory model. *Eur. J. Oper. Res.* **98** 555–562.

Hoffman, P. J. 1960. The paramorphic representation of clinical judgment. *Psych. Bull.* **57** 116–131.

Hughes, H. D. 1917. An interesting seed corn experiment. *Iowa Agriculturist* **17** 424–425, 428.

Hursch, C. J., K. R. Hammond, J. L. Hursch. 1964. Some methodological considerations in multiple-probability studies. *Psych. Rev.* **71** 42–60.

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

REFERENCES

Hyndman, R. J., A. B. Koehler, J. K. Ord, R. D. Snyder. 2008. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer, New York.

Koehler, A. B. 2001. The asymmetry of the sAPE and other comments on the M3-competition. *Internat. J. Forecasting* **17** 570–574.

Lawrence, M., P. Goodwin, M. O'Connor, D. Önkal. 2006. Judgmental forecasting: A review of progress over the last 25 years. *Internat. J. Forecasting* **22** 493–518.

Lawrence, M., M. O'Connor. 1996. Judgement or models: The importance of task differences. *Omega: Internat. J. Management Sci.* **24** 245–254.

Lee, C.-Y., J.-D. Lee, Y. Kim. 2008. Demand forecasting for new technology with a short history in a competitive market: The case of the home networking market in South Korea. *Tech. Forecasting Social Change* **75** 91–106.

Lenk, P., A. Rao. 1990. New models from old: Forecasting product adoption by hierarchical Bayes procedures. *Marketing Sci.* **9** 42–57.

Lindley, D. V. 1983. Reconciliation of probability distributions. *Oper. Res.* **31** 866–880.

Makridakis, S. 1993. Accuracy measures: Theoretical and practical concerns. *Internat. J. Forecasting* **9** 527–529.

McGlothlin, W. H. 1956. Stability of choices among uncertain alternatives. *Amer. J. Psych.* **69** 604–615.

Meehl, P. E. 1957. When shall we use our heads instead of the formula? *J. Counseling Psych.* **4** 268–273.

Mentzer, J. T., C. C. Bienstock. 1998. *Sales Forecasting Management: Understanding the Techniques, Systems and Management of the Sales Forecasting Process*. Sage Publications, Thousand Oaks, CA.

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

Montgomery, A. 1997. Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Sci.* **16** 315–337.

Moriarty, M. 1985. Design features of forecasting systems involving management judgments. *J. Marketing Res.* **22** 353–364.

Morris, P. A. 1974. Decision analysis expert use. *Management Sci.* **20** 1233–1241.

REFERENCES

Morris, P. A. 1983. An axiomatic approach to expert resolution. *Management Sci.* **29** 24–32.

Neelamegham, R., P. Chintagunta. 1999. A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Sci.* **18**(2) 115–136.

Neelamegham, R., P. Chintagunta. 2004. Modeling and forecasting the sales of technology products. *Quant. Marketing Econom.* **2** 195–232.

O'Connor, M., W. Remus, K. Lim. 2005. Improving judgmental forecasts with judgmental bootstrapping and task feedback support. *J. Behav. Decision Making* **18** 247–260.

Pankoff, L. D., H. V. Roberts. 1968. Bayesian synthesis of clinical and statistical prediction. *Psych. Bull.* **80** 762–773.

Pole, A., M. West, J. Harrison. 1994. *Applied Bayesian Forecasting and Time Series Analysis*. Chapman Hall, Boca Raton, FL.

Sanders, N. R., K. B. Manrodt. 1994. Forecasting practices in US corporations: Survey results. *Interfaces* **24** 92–100.

Silver, E. 1965. Bayesian determination of the reorder point of a slow moving item. *Oper. Res.* **13** 989–997.

Theil, H. 1971. *Applied Economic Forecasting*. North-Holland, Amsterdam.

Timmermann, A. 2006. Forecast combinations. G. Elliott, C. Granger, A. Timmermann, eds., *Handbook of Economic Forecasting*. North-Holland, Amsterdam, 135–194.

Tucker, L. R. 1964. A suggested alternative formulation of the developments by Hursch, Hammond and Hursch and by Hammond, Hursch and Todd. *Psych. Rev.* **71** 528–530.

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

Tversky, A., D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Sci.* **185** 1124–1131.

Tyebjee, T. T. 1987. Behavioral biases in new product forecasting. *Internat. J. Forecasting* **3** 393–404.

van Heerde, H. J., C. F. Mela, P. Manchanda. 2004. The dynamic effect of innovation on market structure. *J. Marketing Res.* **41** 166–183.

Webby, R., M. O'Connor. 1996. Judgmental and statistical time series forecasting: A review of the literature. *Internat. J. Forecasting* **12** 91–118.

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

Wright, G., P. Ayton. 1986. The psychology of forecasting. *Futures* **18** 420–439.

Zellner, A. 1986. A tale of forecasting 1001 series: The Bayesian knight strikes again. *Internat. J. Forecasting* **2** 491–494.

Color Figures

Following are full-color renditions of figures that appear in grayscale in the printed paper.

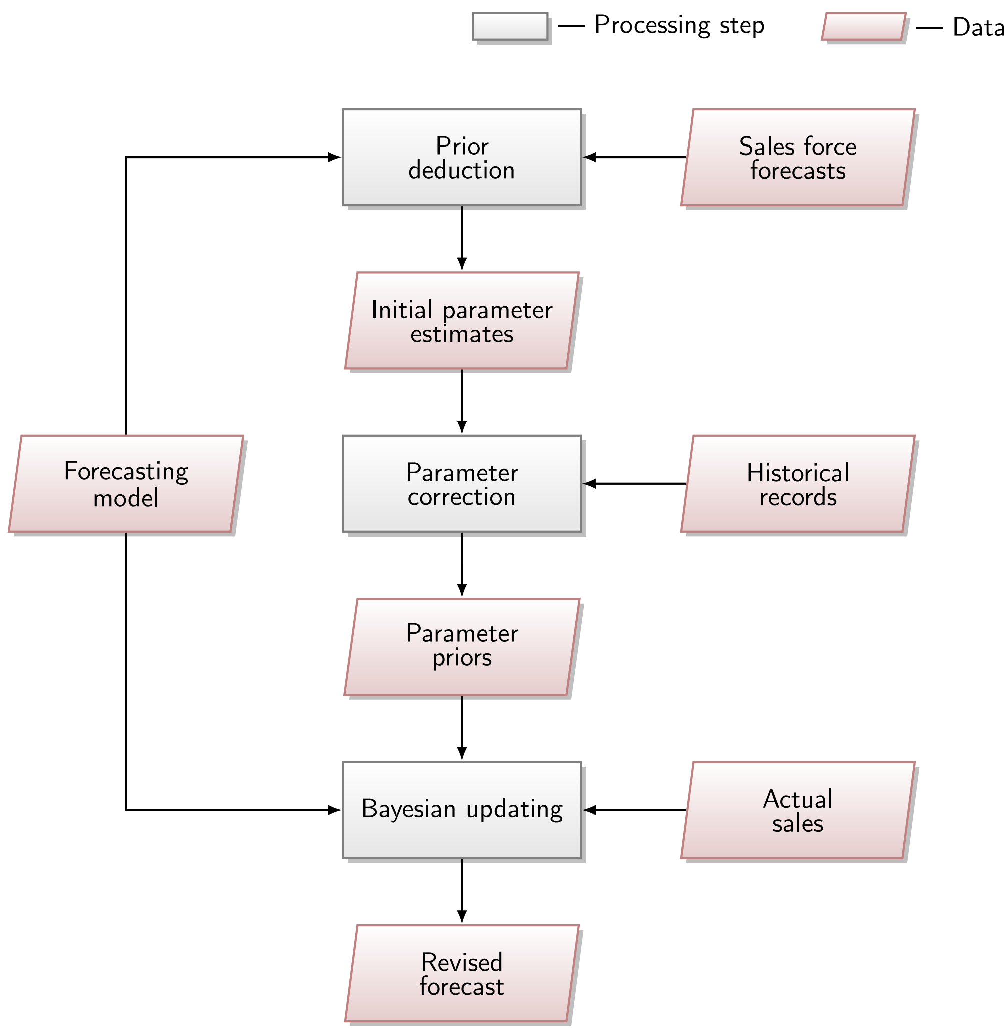

Figure 2: This flowchart provides an overview of the operation of the SLFS.

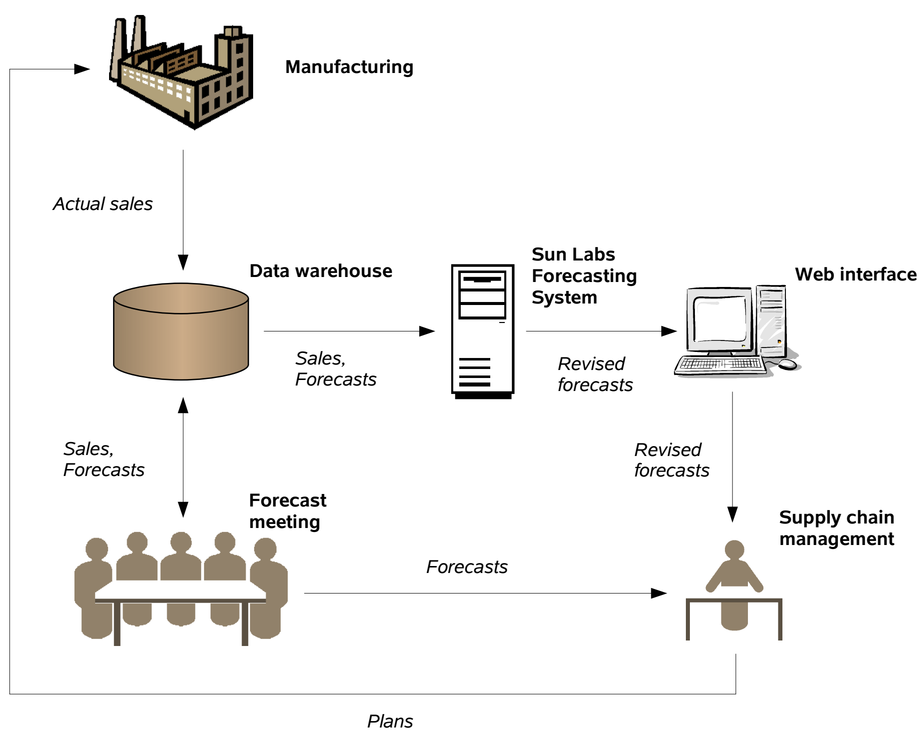

Figure 4: This schematic sketches the way in which the SLFS supports Sun's supply chain management process.

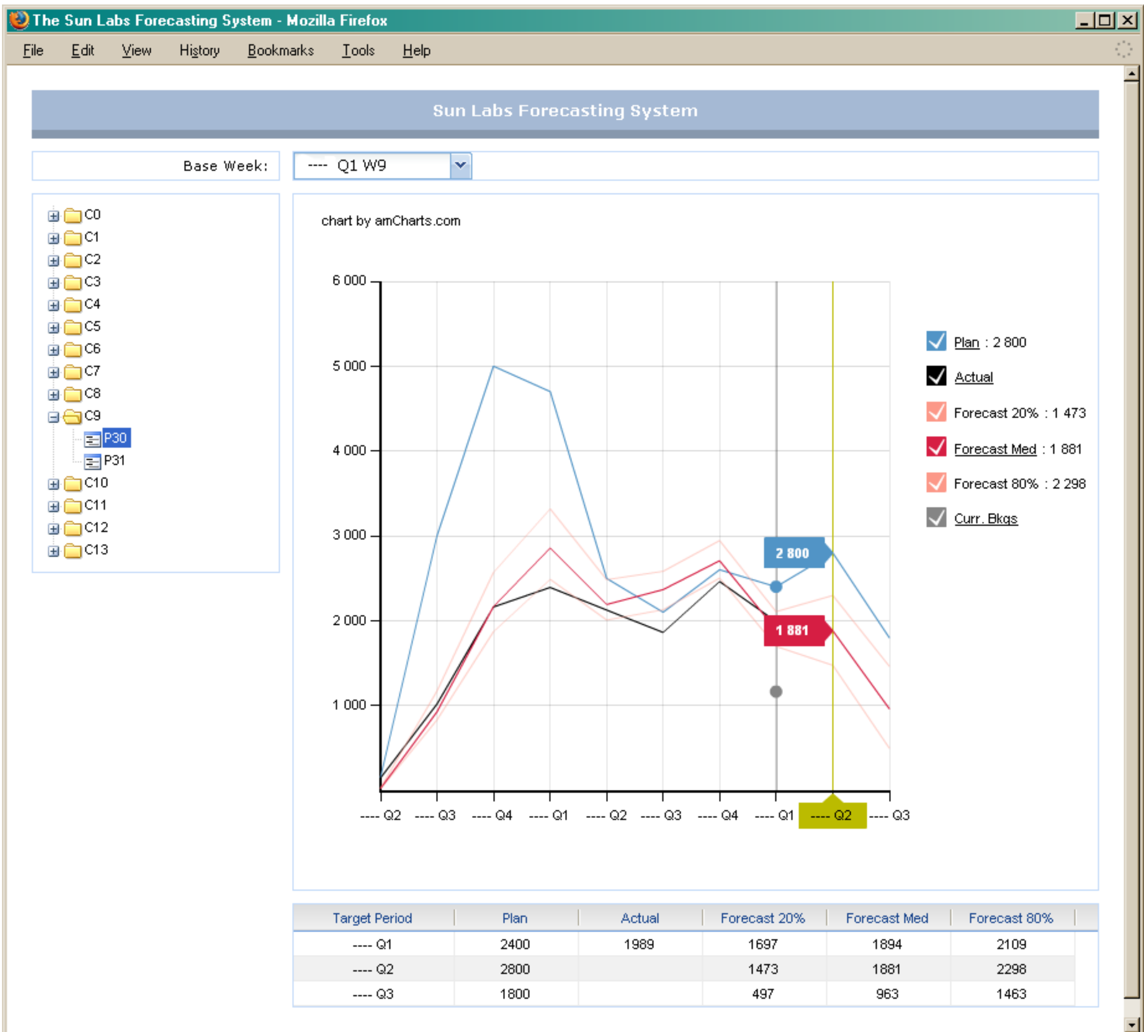

Figure 5: The user interface of the SLFS provides an intuitive web-based graphical view of the system's forecasts. The left side of the page shows the set of forecast products arrayed by category; the top of the page shows a selection of the "base week" in which the forecast was produced, allowing the user to review historical records as desired.