Experiments on Combining Demand Forecasts with Semiconductor Data

Feng Zhang[a], Aliza Heching[b], Sarah Hood[b], Jonathan Hosking[b], Ying-tat Leung[b], Robin Roundy[a,c] and Johnathan Wong[b]

Technical Report 1396, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853, January 2004.

Abstract

In this paper we test several forecast combination methods on industrial demand and forecast data from a semiconductor company. The tested combination methods include four popular methods existing in the literature, and new methods proposed in a companion paper. There are two forecasts to be combined: a time series forecast and a marketing forecast. According to the experiments, combining different demand forecasts does help to improve accuracy. The new combination methods proposed by authors have better and more stable performance than the others. The application of appropriate non-linear transformations to the original demand and forecast data before feeding it into the combination models is strongly recommended.

Index Terms

Demand Forecast, Forecast Combination, Semiconductor

I. Introduction

DEMAND forecasting plays an important role in manufacturing industries and drives many core business processes. In the semiconductor industry forecasting is even more critical than in other manufacturing industries. Business cycles are volatile, price pressure is immense, and competition is unrelenting. Again and again the top management is surprised by unexpected rapid market growth and downturns. Some of the most important and difficult decisions made by semiconductor companies are based on demand forecasts. Examples include production planning, tactical marketing, manpower planning and capacity expansion. An effective forecasting system is critical to the improvement of a manufacturer's competitiveness, revenue and profit.

However, the high volatility of demand in the semiconductor industry makes forecasting challenging. Moreover, it is believed in the industry that forecast accuracy is actually getting worse. According to the survey conducted by Roundy [16], semiconductor manufacturing companies express a fairly low degree of satisfaction with current forecast errors, and an even lower degree of satisfaction with the statistical forecasting methods they are using. The most important reasons for this may be an increasingly complex business environment, increased market fragmentation, shortened product life cycles, and rapidly-changing technology.

This paper addresses an important opportunity to improve forecast accuracy - to statistically combine multiple demand forecasts. In practice, numerous forecasts, plus expert opinion, are available to decision makers. In such a situation, it is difficult to decide which forecasting model or expert will provide the best forecasts. One motivation for combining forecasts has been to avoid the *a priori* choice of a single

[a] Operations Research and Industrial Engineering, Cornell University
[b] IBM Corporation
[c] Corresponding author

forecasting method. Different forecasting methods often model different aspects of the business and make use of different types of information, and are based on different assumptions. Thus, it would be naturally desirable to combine forecasts from time-series, judgemental, and econometric methods and to use data from independent sources such as consumers, producers, retailers, and experts. Also, a combined forecast can function as a risk pooling device, just as investors create diversified portfolios to reduce risk. Combined forecasts are believed to have a smaller risk of an extremely large error than individual forecasts. In both academics and industry, combining forecasts is considered especially useful if the individual forecasting methods are different, and if they draw upon different sources of information.

In general, following forecasts are available to managers in semiconductor companies.

- Time series forecasts, based on statistical models.
- Judgemental forecasts from different sources, such as Marketing, Sales or third party experts.
- Forecasts provided by customers.
- Firm orders placed well in advance of delivery dates.
- Econometric forecasts, normally at the product family level.

Although the methods described here clearly have applications in other domains, we currently focus on short-term demand forecasts which are used to drive production and to support other tactical decisions. These forecasts are typically made at the part level, and usually have planning horizons of 1-6 months into future.

The paper is organized as follows. In Section II we briefly discuss the forecast combination methods tested in the experiments. A summary of the characteristics of the industrial data set, and the results of our computational experiments, are given in Section III. The computational results are done in three phases, and they are followed by our recommendations. In Section IV we summarize and point out potential future research directions .

II. Combination Methods

In [17] we propose a forecast combination framework and derive four different combinational methods by using different weight estimation and forecast selection techniques. Only a brief description is provided here. Interested readers are referred to [17] for more details.

We assume that

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

where $\mathbf{y}$ is an $n \times 1$ vector of historical demand in periods $1 \ldots n$, $\mathbf{X}$ is an $n \times k$ historical forecast matrix, β is a $k \times 1$ vector of unknown weights and ϵ is an $n \times 1$ vector of errors with distribution $N(\mathbf{0}, \sigma^2\mathbf{I})$. A set of linear constraints is applied when we estimate β, namely $\mathbf{0} \leq \beta \leq \mathbf{e}$ and $1 - \delta \leq \mathbf{e}'\beta \leq 1 + \delta$, where δ is set to be $1/4$ and $\mathbf{e}$ is a vector of ones. The use of $\delta > 0$ accommodates a limited amount of bias in the individual forecasts. An indicator function $q(\beta)$ is used to represent these constraints. It takes value of 1 when all constraints are satisfied, and 0 otherwise.

The combining weights β are estimated in a Bayesian approach. The priors are

$$p(\beta|\sigma) \propto N(\mu, \sigma^2\mathbf{V})q(\beta)$$

and

$$p(\sigma) \propto \sigma^{-1},$$

where $N(\mu, \sigma^2\mathbf{V})$ is a multi-variate normal density with mean $\mu = [\frac{1}{k} \dots \frac{1}{k}]'$ and covariance matrix $\sigma^2\mathbf{V} = \sigma^2 k^2 (\mathbf{X}'\mathbf{X})^{-1}$. Applying Bayes theorem and integrating over σ, the posterior distribution for β is $p(\beta|\mathbf{y})$, a multi-variate t distribution truncated to $\{\beta : q(\beta) = 1\}$ (see [17] for a derivation). Two estimation techniques are proposed. One is to take the posterior mean of β, which requires Monte Carlo integration because of the impact of of $q(\beta)$ on $p(\beta|\mathbf{y})$. The other one is to get a Generalized Maximum Likelihood Estimate of β based on $p(\beta|\mathbf{y})$, which is essentially a constrained quadratic minimization problem. We refer to these two techniques as Epost and GMLE, respectively.

In addition to the estimation of β, we also propose two approaches to model selection, which in this context means deciding which individual forecasts to use from a set of available forecasts. This is not considered by most combination methods in the literature. The basic idea of the first selection approach is to find a best from a set of sub-models whose performance will deteriorate more than a threshold when any forecast in that sub-model is dropped. We call it the Economic Significance test. The second one is Bayesian Model Averaging (BMA), which takes an average of all sub-models, weighted by the corresponding posterior sub-model probabilities.

Besides the above methods proposed by the authors, four other existing combination methods are also considered in the experiments. There exist many methods in the literature, ranging from the robust simple average to approaches that are far more theoretically complex, such as state-space methods that attempt to model non-stationarity in the combining weights. According to an extensive survey by Clement [9], simple combination methods often work reasonably well relative to more complex combination. The four methods we choose are simple to understand and implement. All adopt the linear formulation where a vector, $\mathbf{f}$, of k individual forecasts are combined via a linear weighting vector $\mathbf{w}$ to obtain a combined forecast, as $\mathbf{f}_c = \mathbf{w}'\mathbf{f}$.

- Simple Average: Assign equal weight to each individual forecast. This approach has the virtues of simplicity and robustness. It has consistently been the choice of many researchers. Many studies provide strong support for this method (see Clement [9]). Gunter [12] identified analytically the conditions under which the Simple Average outperforms Minimum Variance and OLS, which will be discussed later. A possible answer to the success of Simple Average may rely on the instability of the combining weights, which results from unsystematic changes over time in the covariance matrix of individual forecast errors.
- Outperformance Probabilities: This method is initially proposed by Bunn [7]. By this method, each individual weight is an estimate of the probability that its respective individual forecast performs best on the next occasion. Each probability is estimated as the fraction of occurrences in which the respective individual forecast has been the best in the past. This is a robust, nonparametric method, which performs well when there is relatively little historical data.
- Minimum Variance: The combining weights are calculated in order to minimize the variance of the error of the combined forecast, assuming that each individual forecast is unbiased. Specifically, the weight vector $\mathbf{w}$ is determined by
$$\mathbf{w} = \frac{\mathbf{S}^{-1}\mathbf{e}}{\mathbf{e}'\mathbf{S}^{-1}\mathbf{e}}$$
where $\mathbf{e}$ is a vector of ones and $\mathbf{S}$ is the covariance matrix of individual forecast errors. As $\mathbf{S}$ is generally unknown in practice, this method requires $\mathbf{S}$ to be properly estimated. Bates and Granger [4] suggested five procedures to estimate $\mathbf{S}$. In a finite sample of typical size, sampling error and collinearity among individual forecasts contaminate the estimate of combining weights. Thus, while one hopes to reduce out-of-sample forecast mean squared error (MSE) by combination, there is no guarantee that this will happen in practice.
- Ordinary Least Squares (OLS): In this method the individual forecasts are used as regressors in an ordinary least squares (OLS) regression with a constant term. Granger and Ramanathan [11]

TABLE I
DESCRIPTIVE STATISTICS FOR DEMAND AND FORECASTS OF 3 PRODUCT FAMILIES

Description	Statistics	Range across parts		
		Family 1	Family 2	Family 3
Volatility of demand	std(Dem)/mean(Dem)	[0.57 5.39]	[0.61 4]	[0.88 4.96]
Ability to forecast well	correlation(Dem,TS)	[-0.79 0.73]	[-0.48 0.83]	[-0.52 0.56]
	correlation(Dem,Mkt)	[-0.75 1]	[-0.28 1]	[-0.37 0.98]
Multicollinearity among forecasts	correlation(TS,Mkt)	[-0.56 0.98]	[-0.82 0.86]	[-0.59 0.70]
Forecast variance diversity	std(TS)/std(Mkt)	[0.0016 7.74]	[0.008 12.84]	[0.004 1.86]
Forecast error correlation	correlation(e_{TS},e_{Mkt})	[-0.43 0.94]	[-0.69 0.91]	[-0.18 0.77]
Error variance diversity	std(e_{TS})/std(e_{Mkt})	[0.0019 27.98]	[0.037 10.45]	[0.009 5.06]

showed that if the individual forecasts are biased, this method is better than the Minimum Variance method. Granger and Ramanathan's suggestion has been discussed and contested theoretically (see Clement [8], Bordley [5]) and empirically (see Holden and Peel [13], Aksu and Gunter [1]). Recently, MacDonald and Marsh [14] reported that the presence of substantial biases in individual forecasts led them to use OLS regression to combine exchange rate forecasts.

III. EXPERIMENTS ON INDUSTRIAL DATA

A. *Summary of the Industrial Data*

The industrial data was obtained from a large semiconductor manufacturing company with confidential information disguised. The time span is from Jan. 2000 to Dec. 2002. There are 3 different product families. Hereafter they are referred as Product Family 1, 2 and 3. Roughly there are 230 different parts in Product Family 1, 80 in Product Family 2 and 45 in Product Family 3. Note that not every part exists throughout the 3 year time span. Some entered the market after Jan. 2000 and some were discontinued before Dec. 2002. For each part, historical monthly demand and two historical forecasts were made available to us. One forecast, called TS, is generated by a statistical time series model, and is consequently a function of historical demands. The other one is a marketing forecast (called Marketing). It is based on customer and sales representative forecasts, with adjustments made by marketing executives. Both forecasts were provided with forecast lags of $1, 2, 3, 4, 5$ and 6 months into the future. Unless otherwise noted, data given in this paper refers to an average over all forecast lags (or, in the case of Table I, the union over all forecast lags).

Table I provides some descriptive statistics on the diversity of the demand and forecast data. These statistics include the volatility of the demand, the predictive ability of the TS and Marketing forecasts, the multicollinearity of the TS and Marketing forecasts, and several other interesting diversity measures. Table I indicates that for the industrial data in the study, these measures span wide ranges. As expected, demand volatility is high. Neither forecast is able to predict demand as well as one would like. The relationship between the TS and Marketing forecasts is complicated.

Some combination methods, such as OLS, require a minimum amount of historical data to generate a forecast. In order to make them start to generate combined forecasts from the first time period, we assume that these methods use equal combining weights whenever there is not sufficient historical to make an estimate. This is reasonable because equal weights are the prior belief about β. Once there is sufficient historical data, these methods start to create their own estimates.

B. Accuracy Measurement and Transformation

In our experiments, two forecast accuracy measurements are adopted. The first one is Mean Absolute Error (MAE), defined as

$$\frac{1}{I}\sum_{i=1}^{I}\sum_{t=1}^{T}\frac{|A_{it}-F_{it}|}{T_i}.$$

I is the total number of parts. T_i is the total number of time periods for part i. A_{it} and F_{it} are actual demand and forecasted demand in time period t for part i, respectively. Because MAE sums forecast errors directly, it is dominated by the forecast accuracy for a few parts with high demand volume. To enhance intuition and protect confidentiality, the ratio of MAE and the average demand per period is actually reported. The average demand per period is computed using the same formula used for MAE, with all forecasts equal to zero.

The second forecast accuracy measurement is Geometric Mean of Relative Absolute Error (GMRAE), suggested by Armstrong and Collopy [2]. They evaluate forecast accuracy measures on reliability, construct validity, sensitivity to small changes, protection against outliers and relationship to decision making, and recommend GMRAE. GMRAE is calculated as

$$(\prod_{i=1}^{I} RAE_i)^{\frac{1}{I}}$$

where

$$RAE_i = \frac{\sum_{t=1}^{T}|A_{it}-F_{it}|}{\sum_{t=1}^{T}|A_{it}-G_{it}|}.$$

G_{it} is another forecast for the demand of part i in period t, to which F_{it} is compared. In the following experiments, the Time Series (TS) forecast plays the role of G_{it}. GMRAE is a unitless measure which, in contrast with MAE, assigns equal weight to all parts.

The industrial data in this study has a high degree of skewness, a lot of demands and forecasts that are equal to zero, and other anomalies. Statistical research on linear regression indicates that a non-linear transformation of the data may eliminate lack of fit problems. Box and Cox [6] propose a systematic method for selecting transformations. They suggested using the family of transformations

$$y^{(\lambda)} = \begin{cases} (y^{\lambda}-1)/\lambda & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases}$$

and choosing λ by maximum likelihood.

However, the maximum likelihood method proposed by Box and Cox is not suitable in the current experiment because they apply transformation only to the dependent variable. In this experiment, the independent variables (forecasts) are forecasts of the dependent variable (demand). They are directly comparable numbers. Transformation of the dependent variable alone would destroy this relationship and make the constraints on the combining weights inappropriate.

Since there are many zero demands and forecasts, the transformation we adopt in this work is

$$y^{(p)} = y^{p} \text{ or } y^{(\delta)} = \log(y+\delta),$$

which is applied to both dependent and independent variables. As is pointed out by relevant statistical research, picking a transformation is often a matter of trial and error. Different transformations are tried

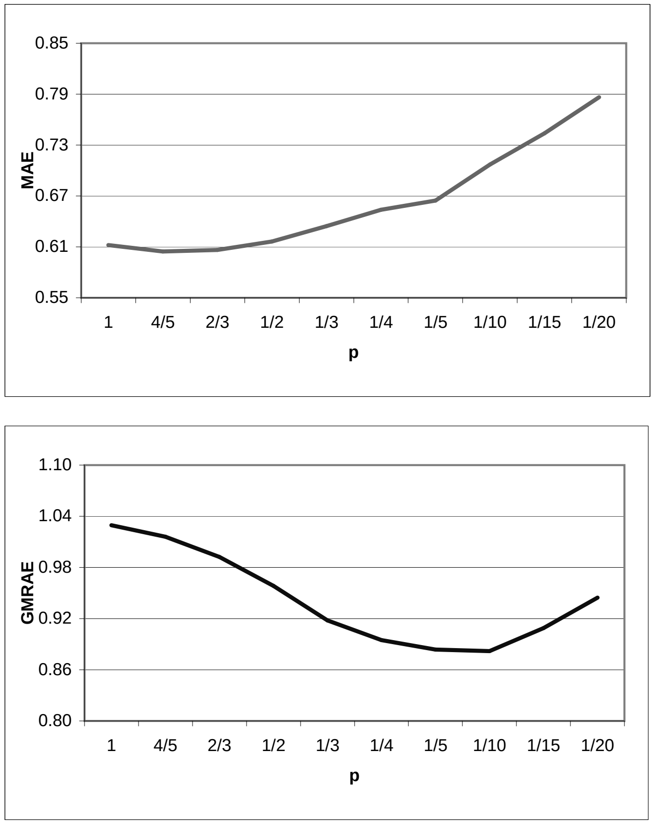

Fig. 1. MAE and GMRAE For Different Power Transformations

until one is found for which ”optimal” performance is obtained [1]. In the current experiments, p is set to be 1, $4/5$, $2/3$, $1/2$, $1/3$, $1/4$, $1/5$, $1/10$, $1/15$ and $1/20$, and δ takes value of 1, 25, 50 and 100. For this data, the power transformation consistently works better than the log transformation.

Figure 1 illustrates how MAE and GMRAE change with respect to the power p in product family 1. The other two families are similar. Both MAE and GMRAE are quasi-convex in p. However, MAE and GMRAE are minimized at different values of p. Which transformation should be applied? As we mentioned before, MAE is essentially dominated by a few parts with high demand volume, while GMRAE is determined by a large number of parts with low demand volume. Consequently each family is divided into two sub-families, by the demand volume. All power transformations are re-tested on the 6 sub-families. MAE and GMRAE retain their quasi-convexity in p. In every case there is a single value of p that approximately minimizes both error measures simultaneously. Figure 2 shows MAE and GMRAE for the 2 sub-families of family 1. The ”optimal” power transformation for each sub-family is summarized in Table II

C. Results using the Standard Approach

In this section the standard approach is tested on the industrial data. According to the standard approach, for a given category of parts, we apply a single nonlinear transformation to the historical forecast and demand data for each part in the category (sub-famly). The categories are defined so that a single parameter

[1]The transformation process works as follows. First historical demands and historical forecasts are transformed. The combination methods estimate combining weights using transformed data. Then, the current forecasts are transformed and combined using these weights. The inverse transformation is applied to this number, to generate the combined forecast in the original context.

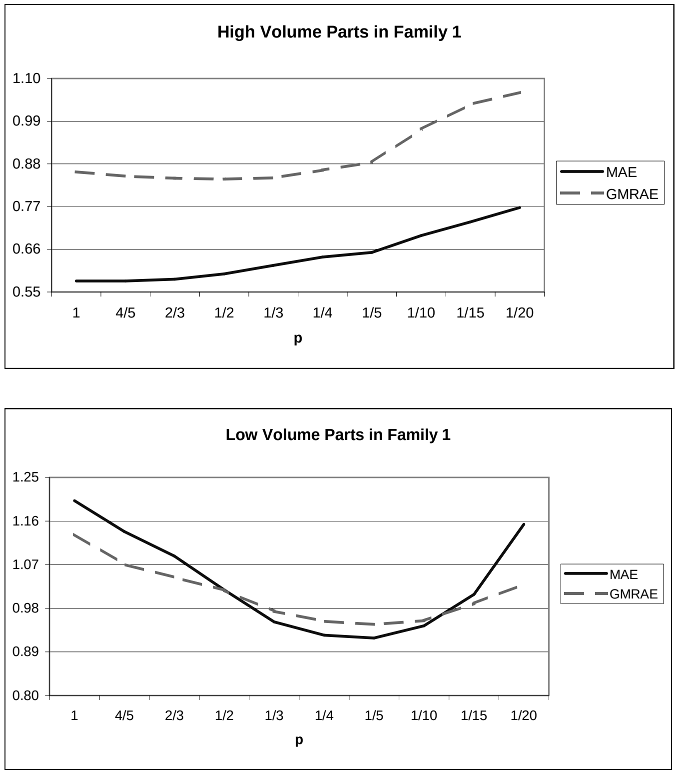

Fig. 2. MAE and GMRAE For High and Low Volume Parts

TABLE II
OPTIMAL POWER TRANSFORMATION FOR EACH SUB-FAMILY

	Family 1		Family 2		Family 3	
	High	Low	High	Low	High	Low
Approx. # of parts	10	220	10	70	10	35
Optimal p	2/3	1/5	1/2	1/5	1/4	1/5

approximately optimizes both the MAE and GMRAE of the combined forecast, as we discussed in section III-B. To be compatible with corporate business processes, we choose to define categories using traditional part families and demand volume. After the data has been transformed, for each part and each forecast lag, we estimate combining weights based on historical demand and forecast data. (Later on the pooling approach will be discussed, in which a group of parts is pooled together to generate a single set of combining weights, which applied to each of the parts in that group.) Tables III, IV and V summarize the results of applying the standard approach to each product family. See the footnote on Table III for the notation used.

All four of the different combination methods proposed by the authors exhibit similar performance. We only report results for Epost(BMA), which seems to be the best method overall, albeit by an insignificant margin. Simple Average (SA) is included too. Outperformance works about as well as SA, and is omitted from the tables. OLS and Minimum Variance are omitted too, due to their unsatisfactory performance. A few observations are worth mentioning.

1) Combination of forecasts does improve accuracy. For GMRAE, combined forecasts dominate individual ones in all sub-families. The improvement is significant, ranging from 13% to 20%. In terms of MAE, combined forecasts beat individual ones in 4 out of 6 sub-families, by 13%-20%. For high-demand parts in families 1 and 2 (approximately 20 out of 325 parts), the marketing forecast

TABLE III

RESULTS USING THE STANDARD APPROACH FOR PRODUCT FAMILY 1

		MAE				GMRAE			
		E(B)[a]	SA	TS[b]	MK[c]	E(B)	SA	TS	MK
High Demand	LT[d] =1	0.48	0.48	0.58	0.53	0.84	0.81	1.00	0.95
	LT =2	0.54	0.52	0.63	0.53	0.79	0.76	1.00	0.79
	LT =3	0.56	0.58	0.68	0.57	0.80	0.79	1.00	0.80
	LT =4	0.55	0.57	0.78	0.57	0.79	0.79	1.00	0.90
	LT =5	0.64	0.64	0.80	0.64	0.86	0.86	1.00	1.02
	LT =6	0.70	0.69	0.82	0.68	0.93	0.91	1.00	1.09
	Avg.[e]	0.58	0.58	0.72	0.59	0.84	0.82	1.00	0.93
Low Demand	LT =1	0.83	0.84	1.04	1.29	0.82	0.87	1.00	0.97
	LT =2	0.81	0.80	1.05	1.04	0.83	0.83	1.00	0.98
	LT =3	0.85	0.85	1.08	1.21	0.86	0.88	1.00	1.15
	LT =4	0.90	0.92	1.13	1.85	0.88	0.94	1.00	1.35
	LT =5	0.98	1.02	1.15	2.45	0.93	1.00	1.00	1.51
	LT =6	1.04	1.10	1.16	3.16	0.93	1.04	1.00	1.66
	Avg.	0.90	0.92	1.10	1.84	0.88	0.93	1.00	1.27

[a]The Epost(BMA) combined forecast
[b]Time Series forecasts
[c]Marketing forecasts
[d]Forecast lag in months
[e]Average over 6 different forecast lags

TABLE IV

RESULTS USING THE STANDARD APPROACH FOR PRODUCT FAMILY 2

		MAE				GMRAE			
		E(B)	SA	TS	MK	E(B)	SA	TS	MK
High Demand	LT =1	0.32	0.32	0.48	0.32	0.78	0.78	1.00	0.77
	LT =2	0.36	0.37	0.55	0.40	0.81	0.83	1.00	0.89
	LT =3	0.46	0.47	0.57	0.43	0.84	0.88	1.00	0.92
	LT =4	0.42	0.48	0.64	0.42	0.84	0.88	1.00	0.92
	LT =5	0.45	0.53	0.70	0.44	0.80	0.85	1.00	0.94
	LT =6	0.49	0.56	0.73	0.47	0.79	0.83	1.00	0.89
	Avg.	0.42	0.46	0.61	0.41	0.81	0.84	1.00	0.90
Low Demand	LT =1	0.77	0.75	0.92	0.89	0.80	0.85	1.00	0.99
	LT =2	0.82	0.84	1.00	1.10	0.87	0.94	1.00	1.31
	LT =3	0.97	0.96	1.13	1.39	0.89	0.96	1.00	1.40
	LT =4	0.93	1.00	1.15	1.43	0.86	0.98	1.00	1.45
	LT =5	0.95	1.03	1.19	1.51	0.85	0.97	1.00	1.46
	LT =6	0.98	1.06	1.23	1.69	0.86	0.99	1.00	1.56
	Avg.	0.90	0.94	1.10	1.34	0.86	0.95	1.00	1.36

is much better than the time series forecast[2]. The combined forecast relies heavily of the marketing forecast. It's MAE as very close to that of the marketing forecast, and its GMRAE is much better.

2) Epost(BMA) is the winner. In one sub-family, high demand parts of family 1, SA is slightly better than Epost(BMA). In the other 5 sub-families, Epost(BMA) is consistently better.
3) The difficulty of forecasting in this domain is illustrated by the MAE. Because of the normalization that we used, the MAE indicates that the average forecast error is usually between 35% and 150% of the average demand.

[2]The Marketing forecast is generated by a marketing organization. In practice one would expect them to devote more time to high demand parts.

TABLE V

RESULTS USING THE STANDARD APPROACH FOR PRODUCT FAMILY 3

		MAE				GMRAE			
		E(B)	SA	TS	MK	E(B)	SA	TS	MK
High Demand	LT =1	0.82	0.90	1.10	1.08	0.70	0.73	1.00	0.78
	LT =2	0.90	0.92	1.06	0.97	0.81	0.82	1.00	0.86
	LT =3	0.92	0.95	1.04	0.98	0.87	0.90	1.00	0.97
	LT =4	0.95	0.99	1.12	1.13	0.82	0.85	1.00	0.94
	LT =5	0.93	1.00	1.13	1.26	0.79	0.85	1.00	0.98
	LT =6	0.98	1.03	1.12	1.45	0.84	0.88	1.00	1.07
	Avg.	0.92	0.97	1.09	1.05	0.80	0.84	1.00	0.93
Low Demand	LT =1	0.98	1.01	1.24	2.78	0.86	0.97	1.00	1.42
	LT =2	0.87	0.90	1.11	1.06	0.74	0.78	1.00	0.84
	LT =3	0.93	0.99	1.14	1.09	0.78	0.84	1.00	0.93
	LT =4	0.99	1.06	1.16	1.23	0.80	0.87	1.00	1.03
	LT =5	0.97	1.06	1.14	1.31	0.82	0.90	1.00	1.08
	LT =6	0.98	1.06	1.14	1.36	0.83	0.89	1.00	1.14
	Avg.	0.95	1.01	1.16	1.47	0.81	0.88	1.00	1.07

D. The Pooling Approach

As we mentioned before, we also test a pooling approach on the industrial data set. The pooling approach works as follows. A group of parts is pooled together. For each forecast lag and each time period, a single set of combining weights is estimated based on all available historical demand and forecast data from items in that group. Subsequently, for each forecast lag, identical weights will be used to generate a combined forecast for each of the parts in the group. One benefit of pooling is the ability to generate combined forecasts for ”young” parts, which have very limited historical data. Other potential advantages include improved statistical power and a reduction in computational time. This is traded off against a potential loss of information, because different parts in the group might be inherently different.

Ideally one wants to pool parts with similar characteristics, in terms of demand and forecasts. Parts in each sub-family are typically similar in technology and function. Therefore, pooling by sub-family seems to be a good starting point. The parts in each product family have already been grouped into two sub-families by demand volume. We use the resulting 6 groups of parts as a basis for data pooling, and refer to this approach as pooling by demand. Table VI summarizes performance of combination methods and individual forecasts, with and without pooling by demand. To simplify the table, only the average performance across six different forecast lags for each method is reported. Again, only one of our combination methods, Epost(BMA), is included (E(B) in the table). Pooling has no impact on the performance of Simple Average (SA) and individual forecasts. Outperformance is not reported because it is dominated by SA. The performance of OLS and Minimum Variance without pooling are both much worse than they are with pooling by demand, and are therefore omitted.

According to Table VI, pooling by demand is unable to substantially improve the accuracy of our combination methods. More precisely, MAE does not improve with pooling, and sometimes it gets significantly worse. In terms of GMRAE, pooling seems to be hurtful for high volume parts and helpful for low volume ones. Another observation is that the performance of OLS is dramatically improved by the pooling approach. It becomes comparable with that of Epost(BMA), but only in terms of GMRAE and only for low volume parts. This is consistent with the observation in experiments on simulated data, OLS requires a lot of data to calibrate (see [17]).

In summary, two simple rules can be abstracted from Table VI.

1) For all high volume sub-families, use Epost(BMA) without pooling.
2) For all low volume sub-families:

TABLE VI

COMPARISON OF STANDARD AND POOLING BY DEMAND APPROACHES FOR ALL FAMILIES

		Family 1		Family 2		Family 3	
		High	Low	High	Low	High	Low
MAE	E(B)[a]	0.58	0.90	0.42	0.90	0.92	0.95
	E(B),PD[b]	0.60	0.94	0.53	0.90	0.91	0.97
	SA	0.58	0.92	0.46	0.94	0.97	1.01
	OLS,PD[c]	0.62	0.97	0.55	0.96	0.96	1.00
	MinVar,PD[d]	0.60	0.92	0.46	0.95	1.02	1.02
	TS	0.72	1.10	0.61	1.10	1.09	1.16
	MK	0.59	1.84	0.41	1.34	1.15	1.47
GMRAE	E(B)	0.84	0.88	0.81	0.86	0.80	0.81
	E(B),PD	0.84	0.86	0.84	0.82	0.81	0.78
	SA	0.82	0.93	0.84	0.95	0.84	0.88
	OLS,PD	0.86	0.86	0.84	0.82	0.87	0.77
	MinVar,PD	0.85	0.91	0.85	0.91	0.91	0.88
	TS	1.00	1.00	1.00	1.00	1.00	1.00
	MK	0.93	1.27	0.90	1.36	0.93	1.07

[a]Epost(BMA) without pooling
[b]Epost(BMA) with pooling by demand
[c]Ordinary Least Squares regression, with pooling by demand
[d]Minimum Variance, with pooling by demand

a) Use Epost(BMA) without pooling, if MAE is the more appropriate accuracy measurement.
b) Use either Epost(BMA) or OLS, with pooling by demand, if GMRAE is the more appropriate accuracy measurement.

In most cases we believe that EPost(BMA) without pooling should be used. The added complexity of using different methods for different part types will dominate the 2%-4% gain in GMRAE for low-volume parts obtained by using the methods of this section, especially when one takes into account that for these parts MAE will suffer by 0%-4%.

E. Product Life Cycle Phases

Another dimension to be explored in the industrial data experiments is product life cycle. The idea is to pool parts by demand and product life cycle phase. The semiconductor manufacturing industry is one of the most dynamic sectors of the world economy. Semiconductor products often have a significantly shorter life span than products in other industries. Increases in speed, reduction in feature size and supply voltage, and changes in packaging technologies occurring with increasing frequency. Based on the 3 years of data that we have, the typical life span of a product ranges from 6 to 30 months and on average is less than 24 months. Most parts go through several life cycle phases corresponding to changes in sales volume. There are five common life cycle phases: introduction, growth, maturity (saturation), declining, and phase-out. Demand and price are typically different for different life cycle phases. For example, the price is high and demand is low but increasing in the introduction phase. Price is low and demand is relatively stable and high in the maturity phase. According to a survey conducted by Roundy [16], semiconductor companies tend to use different forecasting methods in different life cycle phases.

It is quite natural to use life cycle phase in developing and evaluating the performance of forecasting methods. In order to accomplish that, we need the dates of historical life cycle phase transition points for each part. We have inferred these dates from the historical demand data. The methodology adopted is the best known diffusion model in the marketing literature, suggested by Bass [3]. For the industrial data in hand, the Bass model is a helpful but imperfect tool for marking life cycle transition points using historical

TABLE VII

LIFE CYCLE PHASE DEFINITION

Phase	Time Interval	Duration
Ramp-up	Up to T_1	$\frac{1}{(p+q)}\ln[(2+\sqrt{3})\frac{p}{q}]$
Maturity	T_1 to T_2	$\frac{2}{(p+q)}\ln(2+\sqrt{3})$
Declining	T_2 to $2T^*$	$\frac{1}{(p+q)}\ln[\frac{1}{2+\sqrt{3}}\frac{q}{p}]$
Obsolescence	Beyond $2T^*$	-

demand data. Consequently the work described in this section is illustrative, designed to shed light on the magnitude and nature of the benefits that might be attained by using life cycle phase transitions to guide forecast combination techniques, rather than a method that is ready for adoption. Methods to identify and to forecast life cycle transition points, and historical data on the accuracy of those forecasts, are needed before firm recommendations on the use of life cycle phase in forecast combination can be developed. That data is currently not available to us.

The Bass model describes the diffusion process of a new product by the following differential equation:

$$\frac{dN(t)}{dt} = (p + \frac{q}{m}N(t))[m - N(t)]$$

or

$$\frac{dF(t)}{dt} = (p + qF(t))[1 - F(t)] \tag{1}$$

where $N(t)$ is the cumulative demand at time t, m is the ceiling of cumulative demand, p is the coefficient of innovation, q is the coefficient of imitation, and $F(t) = N(t)/m$ is the fraction of adopters who adopt the product by time t. Assuming $F(0) = 0$, simple integration of (1) gives

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + (q/p)e^{-(p+q)t}}$$

Let $f(t)$ be the rate of diffusion at time t. Therefore

$$f(t) = \frac{dF(t)}{dt} = \frac{p(p+q)^2 e^{-(p+q)t}}{(p + qe^{-(p+q)t})^2} \tag{2}$$

The peak time of $f(t)$ is given by (with $q > p > 0$):

$$T^* = -\frac{1}{p+q}\ln(p/q)$$

Note from (2) that the diffusion rate $f(t)$ satisfies $f(0) = f(2T^*) = p$.

Given a Bass diffusion model, another important question remains. How can one identify life cycle phases from this diffusion model? To answer this question, we follow [15] and base the decision on both T^* (the maximizer of $f(t)$) and on the points T_1, T_2 that satisfy $\frac{df(t)}{dt} = 0$ (the inflection points of $F(t)$). Thus we have

$$T_1 = -\frac{1}{(p+q)}\ln[(2+\sqrt{3})\frac{p}{q}]$$

and

$$T_2 = -\frac{1}{(p+q)}\ln[\frac{1}{(2+\sqrt{3})}\frac{p}{q}].$$

The transition time points and duration of each life cycle phase are summarized in Table VII.

In recent years, several estimation procedures have been suggested to estimate the parameters p, q and m of the Bass diffusion model. These include ordinary least squares, maximum likelihood and nonlinear least squares. However each technique has its own shortcoming with respect to providing reliable and accurate forecasts, especially when only a small number of data points is available. A simple simulation based search technique – the Genetic Algorithm (GA) – is applied to minimize the sum of squared error between original and fitted data points for the estimation of p, q and m. As posited by Goldberg [10], GAs are parallel search algorithms that are based on an analogy with Darwin's theory of evolution to converge to a global minimum in a given search space. The inherent nature of GAs ensures that the estimates are robust even with a small number of data points irrespective of the functional form of the objective function. These features of GAs make them as excellent candidate for forecasting applications, with non-linear models such as the Bass model. The MATLAB Generic Algorithm function we use is developed by Michael Gordy at the Federal Reserve System and is available online on his homepage.

Now we are able to dice the six sub-families we have used so far into 18 finer ones according to life cycle phases: rampup, maturity and declining[3]. This gives us new approach to data pooling. We use the phrase ”pooling by demand and life cycle phase” if the parts in one of these 18 sub-families are grouped together when estimating the weights, and the weights are then applied to each of the parts in the sub-family.

Table VIII compares the standard approach, pooling by demand only, and pooling by demand and life cycle phase, for all 18 sub-families. The footnotes to the table explain the notation used. Note that pooling does not have any impact on the Time Series, Marketing and Simple Average forecasts. As we have done heretofore, we only show the most interesting results obtained. Other forecast combination strategies and pooling approaches were also tested.

Four of the 18 sub-families are very interesting. For all low-volume declining parts, and for family 2 high-volume declining parts, the performance of Epost(BMA) with pooling by demand and life cycle phase (E(P),PDL) is dramatically superior to Epost(BMA) without pooling. This is true for both MAE and GMRAE. The relative improvements are in the 10% - 20% range. Moreover, for all low-volume declining parts, using the GMRAE accuracy measure, E(P),PDL is dominated by OLS with pooling by demand (OLS,PD), generating an additional 12% - 20% relative improvement. For Family 2 high-volume declining parts, and for low-volume declining parts using the MAE measure, E(P),PDL and OLS,PD have comparable performance. OLS with pooling by demand and life cycle phase is not competitive.

For the 14 sub-families not discussed in the previous paragraph, the standard approach (Epost(BMA) without pooling) produces satisfactory performance. It is either the best or very close the best for all 14 families, compared to all other combined and individual forecasts. SA is the second best forecast overall. There do exist a few sub-families where either the Time Series or the Marketing forecast dominates. However their performance is not stable.

As was mentioned before, our experiments on product life cycle phase are designed to shed light on the magnitude and nature of the benefits that might be attained by using life cycle phase transitions to guide forecast combination techniques. We are not proposing a method that is ready for adoption. If one is capable of forecasting product life cycle phase transition accurately, and using different forecast combination methods accordingly, more accurate results can probably be obtained for declining parts. Based on the data we have, for declining parts, one would expect improvements in the 10%-30% range. The impact of life cycle phase on forecast combination methods clearly merits further study.

[3] In the experiments, declining is actually the combination of declining and obsolenscence in Table VII. It starts at T_2 and ends when demand dies.

F. Recommendations

For the company that provided us with the data, we recommend the adoption of forecast combination methodology to improve demand forecast accuracy. We recommend that they use Epost(BMA), after applying an appropriate nonlinear transformation to the original demand and forecasts data. It is also recommended that they segment their products by product family and by demand volume, and that for each family-volume pair, they select a different parameter for the non-linear data transformation. The benefit from forecast combination will be 15%-30% on average, depending on how performance is measured. If a simpler method is desired, Simple Average (SA) is the best substitute; the average penalty will be about 4% on average, depending on how performance is measured. It is also recommended that the company not pool data at present. There is strong reason to believe that the identification of life cycle phase transitions and the application of different combination methods according to the life cycle phase will be very beneficial, especially for declining products. However that is premature at present, because the data required to analyze the effectiveness of these methods is not available yet.

Note that the above recommendations are based on data from a single company. The data set does include a fairly wide variety of product lines and markets, at least relative to the semiconductor industry. However data from other companies may exhibit different behaviors. For other companies, we suggest the following general steps for improving demand forecast accuracy by combination.

Step 1: Identify forecasts and other quantifiable sources of information that are potentially useful in demand forecasting. Obtain historical observations for these data streams.

Step 2: Group parts (or products) into product families at a high level according to functionality. For example, ASICs and high-end CPUs might be two of the product families. If a product family has less than 10 parts, statistical significance will suffer. If it is too large then meaningful information will probably be masked. Go through remaining steps for each family.

Step 3: Apply several different forecast combination methods on the historical data. To get a complete view of what is happening, measure performance using both MAE (or another method that weights parts by volume) and GMRAE (or another method that assigns equal weight to all parts). The methods that have been most successful in our work are the methods proposed in [17], Outperformance, Simple Average, and (if you have a substantial amount of historical data or are pooling the data) Ordinary Least Squares.

Step 4: Apply a group of power transformation functions to see if further improvement can be achieved. See Section III-B for details. The recommended range for the power p is between $1/10$ and 1.

Step 5: If the MSE and GMRAE measures of forecast accuracy indicate that different optimal data transformations or different combination methods are recommended, consider partitioning the product family into sub-families, and repeat Steps 3 and 4.

Step 6: Check the impact of pooling on the combination methods, to see if it leads to improved performance.

TABLE VIII

COMPARISON OF STANDARD AND POOLING APPROACHES FOR ALL FAMILIES

		MAE						GMRAE					
		High			Low			High			Low		
		Ramp-up	Maturity	Declining	Ramp-up	Maturity	Declining	Ramp-up	Maturity	Declining	Ramp-up	Maturity	Declining
Fam. 1	E(B)[a]	0.56	0.59	0.44	0.95	0.78	1.07	1.12	0.83	0.78	1.09	0.91	0.80
	E(B),PD[b]	0.60	0.60	0.45	0.98	0.81	1.04	1.17	0.83	0.78	1.08	0.93	0.76
	E(B),PDL[c]	0.63	0.59	0.54	0.97	0.90	0.96	1.22	0.84	0.76	1.06	0.94	0.65
	SA	0.56	0.60	0.44	0.95	0.79	1.15	1.11	0.82	0.77	1.10	0.92	0.86
	OLS,PD[d]	0.70	0.61	0.53	0.98	0.96	0.98	1.22	0.84	0.74	1.05	0.95	0.57
	TS	0.77	0.74	0.44	1.00	0.88	1.65	1.00	1.00	1.00	1.00	1.00	1.00
	MK	0.59	0.59	0.54	2.64	1.41	2.42	1.29	0.91	0.81	1.31	1.07	0.99
Fam. 2	E(B)	0.62	0.32	1.06	0.90	0.84	1.14	0.79	0.85	0.73	1.11	0.94	0.68
	E(B),PD	0.62	0.46	1.01	0.96	0.84	1.09	0.86	0.89	0.70	1.11	0.92	0.62
	E(B),PDL	0.62	0.46	0.93	0.95	0.89	0.89	0.87	0.88	0.64	1.08	0.95	0.50
	SA	0.61	0.37	1.12	0.90	0.85	1.25	0.79	0.87	0.77	1.12	0.97	0.76
	OLS,PD	0.66	0.51	0.93	0.98	0.96	0.94	0.90	0.91	0.63	0.95	0.97	0.41
	TS	0.76	0.52	1.38	0.90	0.91	1.89	1.00	1.00	1.00	1.00	1.00	1.00
	MK	0.69	0.27	1.35	1.39	0.86	1.71	0.78	0.83	0.84	1.37	1.15	1.07
Fam. 3	E(B)	0.95	0.95	0.93	1.01	0.85	1.32	0.93	0.96	0.69	1.08	0.77	0.78
	E(B),PD	0.94	0.95	0.90	1.01	0.92	1.24	0.89	0.99	0.67	1.06	0.80	0.72
	E(B),PDL	0.94	0.92	0.87	1.00	0.93	1.04	0.88	0.98	0.65	1.02	0.79	0.60
	SA	0.95	0.92	1.37	1.02	0.88	1.44	0.95	0.93	0.74	1.12	0.80	0.86
	OLS,PD	1.08	0.97	0.88	1.00	0.99	1.01	0.94	1.11	0.66	1.00	0.81	0.48
	TS	1.02	0.97	1.37	1.03	1.02	1.78	1.00	1.00	1.00	1.00	1.00	1.00
	MK	1.25	1.06	1.00	1.50	1.02	1.87	1.25	1.05	0.73	1.45	0.75	1.01

[a]Epost(BMA) without pooling
[b]Epost(BMA) with pooling by demand
[c]Epost(BMA) with pooling by demand and life cycle phase
[d]Ordinary Least Squares regression, with pooling by demand

IV. Conclusions

The purpose of this paper is explore the possibility of improving demand forecast accuracy by combining different forecasts, using data from a large semiconductor manufacturer. We report results for four commonly used combination methods, and for EPost(BMA) (one of the methods proposed in [17]). We test these combination methods on industrial data for three different product families from a semiconductor manufacturer. Forecast combination leads to significant accuracy improvement for all of the families. It is important to group products into families or sub-families carefully, and to apply appropriate non-linear transformations to original demand and forecasts data before feeding it into the combination models. The idea of pooling different parts together for statistical estimation is examined as well, both by family and demand, and by family, demand and life cycle phase. Among the tested models, Epost(BMA) without pooling is the most accurate and robust method, and is recommended for all product families. The exception is for low volume parts in their declining life cycle phase, where pooling helps, and ordinary least squares is competitive with EPost(BMA). However more research is required to exploit the impact of life cycle phase on forecast combination methods, and we currently lack the data required to do the job properly.

In all of our work we have grouped parts in ways that are compatible with business processes. In this case, that means using traditional product family definitions, demand volume, and life cycle phase. If compatibility with business processes is not a priority, there are many partitioning methods that could be used to group the parts. These groupings could be used for data pooling and/or determining which non-linear transformations are most appropriate.

V. Acknowledgement

The authors are deeply indebted to the corporation that provided us with the data. We also thank the Semiconductor Research Council, the National Science Foundation and IBM for generous financial support, without which this research would not have been possible.

References

[1] C. Aksu and S. I. Gunter, *An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination of forecasts*, International Journal of Forecasting, vol. 8, pp. 27-43, 1992.

[2] J. S. Armstrong and F. Collopy, *Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons*, International Journal of Forecasting, vol. 8, pp. 69-80, 1992.

[3] F. M. Bass, *A New Product Growth Model for Consumer Durables*, Management Science, vol. 15, pp. 215-227, 1969.

[4] J. M. Bates and C. W. J. Granger, *The combination of forecasts*, Operational Research Quarterly, vol. 20, pp. 451-468, 1969.

[5] R. F. Bordley, *Technical note: Linear combination of forecasts with an intercept: A Bayesian approach*, Journal of Forecasting, vol. 5, pp. 243-249, 1986.

[6] G. E. P. Box and D. R. Cox, *An Analysis of Transformations*, Journal of the Royal Statistical Society, Series B, vol. 26, pp. 211-246, 1964.

[7] D. W. Bunn, *A Bayesian approach to the linear combination of forecasts*, Operational Research Quarterly, vol. 26, pp. 325-329, 1975.

[8] R. T. Clement, *Linear constraints and the efficiency of combined forecasts*, Journal of Forecasting, vol. 5, pp. 31-38, 1986.

[9] R. T. Clement, *Combining forecasts: A review and annotated bibliography*, International Journal of Forecasting, vol. 5, pp. 559-583, 1989.

[10] D. E. Goldberg, In *A Simple Genetic Algorithm in Genetic Algorithms in Search Optimization and Machine Learning*, Addison Wesley Longman, 1989, 10-14.

[11] C. W. J. Granger and R. Ramanathan, *Improved methods of forecasting*, Journal of Forecasting, vol. 3, pp. 197-204, 1984.

[12] S. I. Gunter, *Theoretical justification of the efficiency of simple average combinations*, Working Paper, 1990, Temple University, Philadelphia, PA

[13] K. Holden and D. A. Peel, *Unbiased, efficiency and the combination of economic forecasts*, Journal of Forecasting, vol. 8, pp. 175-188, 1989.

[14] R. MacDonald and I. W. Marsh, *Combining exchange rate forecasts: What is the optimal consensus measure?*, Journal of Forecasting, vol. 13, pp. 313-333, 1994.

[15] V. Mahajan, E. Muller and R. K. Srivastava, *Determination of Adopter Categories by Using Innovation Diffusion Models*, Journal of Marketing Research, vol. 27, pp. 37-50, 1990.

[16] R. Roundy, *Report on practices related to demand forecasting for semiconductor products*, Technical Report No. 1293, 2002, School of O.R. and I.E., Cornell University.

[17] F. Zhang and R. Roundy, *Improving Forecast Accuracy by Combination*, Technical Report No. 1395, 2003, School of O.R. and I.E., Cornell University.