Market Sales Forecasting Related to the Semiconductor Manufacturing Industry

Jingyi Guo[a]
Department of Statistics, Rutgers University, New Brunswick, U.S.A.

Keywords: Multiple Regression Analysis, Semiconductor Manufacturing Industry, Market Sales Forecast.

Abstract: Based on the multiple linear regression model, this paper analyzes and compares the effects of the sales forecasting models SMT and CIT on the relevant markets of the US semiconductor manufacturing industry. To be specific, the data is used from the ABCtronics case to conduct a multiple regression analysis on the status of the semiconductor manufacturing industry. According to the results of multiple regression analysis, the proposed model of SMT interns has strong explanatory power than the model previously used at ABCtronics. These results shed light on sales prediction in terms of different relevant variables, which can be implemented to different industries.

1 INTRODUCTION

The production of semiconductor manufacturing industry is closely correlated with the cyclical demand pattern and the complex nature of chip design. As a result, it undergoes an economic challenge because of the excess capacity derived from lower demand, and the high research and development cost along with the increasing fixed cost of building the most advanced facilities for wafer fabrication (G Anderson, Sweeney, Williams, Camm, Cochran, 2011). ABCtronics, one of the companies within the industry, recently has been dealing with issues on improvement of products' quality control tests, plants' downtime, and analysis of customer feedback.

The first problem is associated with chips' quality control tests. ABCtronics has been using Lot Acceptance Testing Method (LATM) for quality control, taking a sample of 25 IC chips from a lot of 500 without replacement. A lot will pass the check if the sample has less than 2 defective chips. The method has shown that on average, every 500 ICs has 2 defective chips, and entails that the probability of producing a defective chip is 0.004. ABCtronics is debating on whether to use Individual Chip Testing System (LCTM) as an alternative, which is a similar statistical method that takes a sample of 25 IC chips from a lot of 500 but with replacement. On this basis, they hope to narrow down the probability of defective production to 0.002. With regards to high downtime and chemical impurities, Mark, the head of QRT, proposes that the plants' average downtime should be reduced to 5 hours. On the contrary, Stuart, the president of the fabrication plant, states that they have to replace their ion implanter because the machine and its subsequent activities follow a uniform distribution instead of a Gamma distribution. Moreover, he and Mark point out that the percentage of chemical impurities per lot follows a beta distribution. According to the policy, it will not be used in the fabrication process if the impurities take up more than 30%, and the process has been working fine.

Another urgent issue is their analysis on customer relationships and prediction of sales figure. For customer feedback, ABCtronics randomly surveyed 40 customers for their 74XX chip family. 32 out of 40 chosen customers rated their products beyond good or satisfactory. Jim is skeptical about the results because the overall spread of the customer feedback rating is very high even though the mean is 56. He holds that redesigning the survey, e.g., taking more samples of customers might help. Furthermore, one highly suggests that the company should use multiple linear regression models for predicting the sales figure (Berenson, Mark, et al, 2012). Therefore, multicollinearity problems can be avoided when dealing with sales figures for various demand scenarios (Maxwell 2000). The rest part of the paper

[a] https://orcid.org/ 0000-0002-9782-0501

Guo, J.
Market Sales Forecasting Related to the Semiconductor Manufacturing Industry.
DOI: 10.5220/0011235800003440
In *Proceedings of the International Conference on Big Data Economy and Digital Management (BDEDM 2022)*, pages 714-719
ISBN: 978-989-758-593-7

is organized as follows. The Sec. 2 will introduce the data origination and statistic principles, Subsequently, the Sec. 3 will display the results of linear regression. Afterwards, the explanation of the results as well as the limitation of the method will be demonstrated in Sec. 4. Eventually, a brief summary is given in Sec. 5.

2 DATA &METHOD

If a sample n is taken without replacement from a finite (small) population of size N in which M has an attribute (and hence N – M do not possess that attribute), the number of sample units (Wang 2021), X possessing that attribute follows a Hyper-geometric distribution with parameters N, M, n. Probability mass function (pmf) is:

$$p(x) = P(X = x) = \frac{(M^{Cx})(n - m^{Cn} - x)}{N^{Cn}} \quad (1)$$

This probability can also be directly obtained using Excel Function-Statistical-Hypgeomdist. If X~B(n, p). i.e., X has Binomial Distribution with parameters n and p, where n = number of trials and p = probability of one success, then, probability mass function (pmf) of X is given by

$$p(x) = P(X = x) = n^{Cx} p^x (1 - p)^{n-x} \quad (2)$$

This probability can also be directly obtained using Excel Function: Statistical, BINOMDIST. To work out the solution, N = 500, n = 25, Number of defectives in the lot = 2, i.e., proportion of defectives in the lot, p = 0.004. Let X represents number of defectives found in a sample of 25. Current system (LATM), one can easily get the data. Based on Eq. (1), X ~ Hyper-geometric with N = 500, M = 2, n = 25. P (lot is accepted) = 0.9976. Proposed system (ICTM), Here X ~ B (25, 0.004), P (lot is accepted) = 0.9047.

Table 1: Table Type Style.

Year	Sales Volume	Market demand	Price per chip	Condition
2004	2.39	297	0.832	0
2005	3.82	332	0.844	1
2006	3.33	195	0.854	0
2007	2.49	182	1.155	1
2008	1.56	93	1.303	0
2009	0.97	98	1.265	0
2010	1.32	198	1.368	1
2011	1.42	188	1.208	0
2012	1.48	285	1.234	1
2013	1.85	264	1.282	1

The variables for regression are summarized in Table. I. For the same lot size and defective level, ICTM would reject the lot more often than LATM. Thus, it is more stringent. Flaw in the current system. For a defective level of 0.4%, an acceptance number of 1, which out of 25 is 4%, is clearly too lenient. Consequently, lots with higher defectives level are likely to get accepted and passed on to the customer.

Primarily, one needs to count the expression of the exponentially distributed, and its probability density function is $f(x)=\lambda e^{-\lambda x}$ ($\lambda > 0$), when $\lambda < 0$, $f(x)=0$. Besides, its distribution function is $F(x)=1-e^{-\lambda x}$. From the picture "probability density function", one can easily know that when x=5, f(x)=4.1% (assuming that x=time before IC chip failure (in years)). On this basis, the expression can be derived as $0.041=\lambda e^{-5\lambda}$. However, at this time, one can easily solve this problem, because the data is not integer. Thus, it should use the distribution function to help us simplify the count process, through two expressions, one can easily know the number of λ instead of using complicated count (Choi 2021). According to the data, one finds that x=5, F(x)=0.2255 (assuming that x=time before IC chip failure), i.e., $0.2255=1-e^{-5\lambda}$. Then put two equations together, one gets that the λ=0.041/0.7745=0.053 (assuming that e=2.7), i.e., the $f(x)=0.053e^{-0.053x}$ and $F(x)=1-e^{-0.053x}$. Besides, Customer PQR systems request that the chips will last more than 6 years, i.e., x≥6. Substituting x=6, the F(x)=0.27, f(x)=3.86%. In this case, Mark is confident that ABCtronics should be able to meet the expectation of the client Customer PQR systems.

They have again started experimenting with their quality control. Circuit module M (CM) has a path where three chips from ABCtronics get connected in a series. Before the new testing process, XYZsoft reported that in a typical lot comprising 20 CMs they are finding three defective items. In most of those cases, they observed that the problem was with our chips. Now, they have put a stricter policy in place. They have now started to calculate the number of nondetective before they encounter a particular number of defectives.

According to the exponentially distribution, the probability of chip failure keeps decreasing as the time period increases (Lee 2017). Based on the cross comparison with the cumulative distribution function of failure time, the graph of the cumulative exponential function also shows a decrease in its slope, meaning that the chance to fail shrinks after year 5, which results in an up to forty-year long-lasting value of IC chips. This is an evident reason why the mark was so confident that PQR systems expectations can be met. The increase in the number of complaints regarding ABCtronics' IC chips started when XYZsoft was required to retrieve the whole lot of

chips for rework and recheck once there was a defective item being detected (Sawyer, Richard, 1982).

In retrospect, the average HIGH signal output was 2.7V, which fell within the tangible range of a good IC chip. However, when the new testing policy is adopted, the average HIGH signal output drops to 2.3v, which suggests a problematic output. The reason could be ABCtronic's method of sampling random 100 chips across lots for the quality test. There is no separate chip testing only for those defective ones to get their values of output, i.e., ABCtronic can have a better idea on how far these chips are away from the standards given by XYZsoft. As a result, it is considered that ABCtronic is over-estimating the output of the chip. In order to improve their estimation, they don't need to return the defective chips. Given N=40, Sum of the observations=2276, mean $\bar{x}$ =2276/40=56.9, the standard derivation=18.98 with the 90% confidence interval (51.965, 61.835), which agrees with Robert's analysis of customer score that they are doing either good or satisficing with his products. Based on statistic analysis, n=61 is minimum sample size required to shrink the margin of error to 4, i.e., analyze the mean customer score with 90% confidence.

3 RESULTS

Let ABCtronics' sales volume, overall market demand, price per chip, and economic condition be denoted by 1, 2, and 3, respectively. ABCtronics was using a simple linear regression model for predicting its sales figure. As a result, the estimated linear regression equation of sales volume on overall market demand is $\hat{Y} = 0.7534 + 0.00614X_1$. From the R^2 Value, it can be concluded that only 28.49% of the variation in the sales volume is explained by this regression model. By using all three available variables, if one runs a multiple linear regression model, model Y will depict the effect of X_1, X_2, and X_3 on Y. The estimated multiple linear regression equation of sales volume on overall market demand, price per chip, and economic condition is given by: $\hat{Y} = 8.8607 - 0.0052X_1 - 5.5.54X_2 - 1.1302X_3$. At 5% level of significance, the p-value for the overall market demand (X1) is more than 0.05 which is insignificant. Ignore X_1, again the regression analysis of Y and X_2, X_3 is performed. $\hat{Y} = 6.4517 - 4.1356X_2 + 0.6062X_3$, for this multiple linear regression model, the calculated value of the adjusted R^2 is 0.7996. In order to check for multicollinearity, i.e., to see if there exists any linear relationship between independent variables, they have computed the variance inflation factor (VIF) between X2 and X3 which is given by $VIF_{X_2,X_3} = \frac{1}{1-RX_2X_3{}^2} = 1.0473$. The computed value of VIF signifies that the modified multiple linear regression model is not suffered by the problem of multicollinearity. The results for different models are listed in Tables II-XIII.

Table 2: Simple linear regression model.

ANOVA	df	SS	MS	F	Significance F
Regression	1.000	2.217	2.217	3.188	0.112
Residual	8.000	5.563	0.695		
Total	9.000	7.780			

Table 3: Simple linear regression model.

Regression Statistics	
Multiple R	0.534
R Square	0.285
Adjusted R Square	0.196
Standard Error	0.834
Observations	10.000

Table 4: Simple linear regression model.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.753	0.779	0.967	0.362	-1.044	2.551	-1.044	2.551
Makert demand	0.006	0.003	1.786	0.112	-0.002	0.014	-0.002	0.10

Table 5: Multiple linear regression model.

Regression Statistics	
Multiple R	0.953
R Square	0.908
Adjusted R Square	0.861
Standard Error	0.346
Observations	10.000

Table 6: Multiple linear regression model.

ANOVA	df	SS	MS	F	Significance F
Regression	3.000	7.061	2.354	19.628	0.002
Residual	6.000	0.719	0.120		
Total	9.000	7.780			

Table 7: Multiple linear regression model.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	8.861	1.348	6.572	0.001	5.562	12.160	5.562	12.160
Makert demand	-0.005	0.003	-2.028	0.089	-0.012	0.001	-0.012	0.001
Price per chip	-5.505	0.881	-6.246	0.001	-7.662	-3.349	-7.662	-3.349
Condition	1.130	0.342	3.304	0.016	0.293	1.967	0.293	1.967

Table 8: Multiple linear regression model(X2,X3 variables).

Regression Statistics	
Multiple R	0.919
R Square	0.844
Adjusted R Square	0.800
Standard Error	0.416
Observations	10.000

Table 9: Multiple linear regression model(X2,X3 variables).

ANOVA	df	SS	MS	F	Significance F
Regression	2.000	6.568	3.284	18.960	0.001
Residual	7.000	1.212	0.173		
Total	9.000	7.780			

Table 10: Multiple linear regression model(X2,X3 variables).

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	6.452	0.766	8.422	0.000	4.640	8.263	4.640	8.263
Price per chip	-4.136	0.680	-6.079	0.001	-5.744	-2.527	-5.744	-2.527
Condition	0.606	0.269	2.250	0.059	-0.031	1.243	-0.031	1.243

Table 11: Test multicollinearity.

Regression Statistics	
Multiple R	0.213
R Square	0.045
Adjusted R Square	-0.074
Standard Error	0.216
Observations	10.000

Table 12: Test multicollinearity.

ANOVA	df	SS	MS	F	Significance F
Regression	1.000	0.018	0.018	0.379	0.555
Residual	8.000	0.374	0.047		
Total	9.000	0.392			

Table 13: Test multicollinearity.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.092	0.097	11.293	0.000	0.869	1.315	0.869	1.315
Condition	0.084	0.137	0.616	0.555	-0.231	0.400	-0.231	0.400

4 DISCUSSIONS

As a matter of fact, the sales prediction model proposed by SMT interns has both advantages and disadvantages. As for disadvantages, the multiple linear regression is more effective and practical to predict or estimate the dependent variable by the optimal combination of several independent variables (Heinrichs, Nina, et al, 2009). Moreover, it also considers the price per chip and economic condition, hence it may be more all-around. Regarding to advantages, the sales prediction model proposed by SMT interns is simpler and easier to count. More importantly, the effect of overall market demand is larger than other two factors, one can deduct that actually price per chips and economics condition both contribute to the overall market demand, when price is high, the demand is low, on the contrary, the price is low, demand is high. Besides, the economic condition is good, the demand is high. Therefore, one can conclude that the overall market demand directly affects the sales volume. In addition, Phil's model divides overall market demand into 3 level, the same as sale volumes, which clearly demonstrate the relationship between them.

Expected sales is about 2.061, thus the expected sales figure for ABCtronics in this year is 2.061. From the results in Tables II-XIII, it can easily know that, when sales volume reach 3 million, total market demand of PCs is over 200 and between 100 and 200, i.e., the chance is 0.1+0.1=0.2.

As the semiconductor manufacturing company, ABCtronic faces an economic challenge for two reasons: (i) cyclical nature of demand and (ii) the high cost associated with research and development. Usually, one uses the way of monitoring some key parameters for deviations to solve "major scrap events" (Prescott, Patricia A 1987, Yang 2019). Moreover, the

other problems are also discussed in memo part. Additionally, it is considered that ABCtronic should pay more attention to the XYZsoft, one of the major clients of ABCtronics, uses IC chips on their personal computers (PCs). Now that, the ABCtronic is the biggest customer and also be the main source of income, ABCtronic should avoid the problems about chips, which cause much compliant from XYZsoft, and let it return some chips to recheck and rework it. From long term, it may do some damage to the stable partnership between ABCtronic and XYZsoft.

5 CONCLUSION

In summary, the sales prediction based on multifactorial linear regression models of ABCtronic is carried out. According to the results, he ABCtronic should put the quality of products at the first position. Based on the analysis, one sees that the whole semiconductor manufacturing industry is facing fierce competition, and every company should face the demand of market whenever it be good or bad, at this time, the innovation or quality will be the most powerful weapon. After all, it means less rework and longer lifetime, which will save a lot of money compared with the large investment in research and development. Overall, these results offer a guideline for sales prediction in terms of multi-variables.

REFERENCES

Berenson, Mark, et al. Basic business statistics: Concepts and applications. Pearson higher education AU, 2012.

Choi, Hyun Il. "Development of Flood Damage Regression Models by Rainfall Identification Reflecting Landscape Features in Gangwon Province, the Republic of Korea." Land 10.2 (2021): 123.

G Anderson, D., Sweeney, D., Williams, T., Camm, J., Cochran, J. 2011. Statistics for Business & Economics, 11th ed. Cengage Learning, Mason.

Heinrichs, Nina, et al. "Die Wirksamkeit ambulanter Psychotherapie der Sozialen Angststörung in einer universitären Ambulanz: Wird die Forschung in die Praxis transportiert?." Zeitschrift für Klinische Psychologie und Psychotherapie 38.3 (2009): 181-193.

Lee, In. "A study of the effect of social shopping deals on online reviews." Industrial Management & Data Systems (2017).

Maxwell, Scott E. "Sample size and multiple regression analysis." Psychological methods 5.4 (2000): 434.

Prescott, Patricia A. "Multiple regression analysis with small samples: Cautions and suggestions." Nursing Research (1987), 36, (2): 130.

Sawyer, Richard. "Sample size and the accuracy of predictions made from multiple regression equations." Journal of Educational Statistics 7.2 (1982): 91-104.

Wang, Kung-Jeng, Pei-Shan Wang, and Phuc Hong Nguyen. "A data-driven optimization model for coagulant dosage decision in industrial wastewater treatment." Computers & Chemical Engineering (2021): 107383.

Yang, Qifan, et al. "Revisiting the relationship between correlation coefficient, confidence level, and sample size." Journal of chemical information and modeling 59.11 (2019): 4602-4612.