Modelling information sharing to improve just-in-time purchasing vendor evaluation

James T. Ding and Rohana J. Karunamuni*

Decisive Knowledge Consulting LLP,
555 California Street,
San Francisco, CA 94104, USA

Department of Mathematical and Statistical Sciences,
University of Alberta,
Edmonton AB T6G 2G1, Canada
E-mail: jtding@dkllp.com
E-mail: R.J.Karunamuni@ualberta.ca
*Corresponding author

Abstract: Vendor evaluation is an important step for a manufacturer's purchasing operation. For firms adopting Just-In-Time Purchasing (JITP) strategy, vendor evaluation is critical in quality improvement and cost reduction. In this paper, we examine the value of information sharing in improving vendor evaluation. We build stochastic models for three different scenarios of information sharing to illustrate the advantages for JITP vendor evaluation under the Assemble-To-Order (ATO) manufacturing environment. We show how integrating statistical and analytical tools with Enterprise Resource Planning (ERP) packages can help achieve efficient and effective managerial decision-making in JITP.

Keywords: assemble-to-order manufacturing; ATO; enterprise resource planning; ERP; information sharing; just-in-time purchasing; JITP; quality control; QC; vendor evaluation.

Reference to this paper should be made as follows: Ding, J.T. and Karunamuni, R.J. (2008) 'Modelling information sharing to improve just-in-time purchasing vendor evaluation', *Int. J. Manufacturing Technology and Management*, Vol. 13, No. 1, pp.30–54.

Biographical notes: James T. Ding is a partner with Decisive Knowledge, a boutique consulting firm specialising in quantitative decision analysis for value chain integration solutions. He has over ten year professional experience in manufacturing industry and consulting. He received a BS in Mathematics, an MS in Statistics and also was a PhD candidate in Management Science and Information Systems. His current research interests are in the areas of Bayesian decision analysis, statistical computing, quality engineering, service industry and supply chain management.

Rohana J. Karunamuni is a Professor in Statistics, Department of Mathematical and Statistical Sciences at University of Alberta. His current research interests are in decision theory, non-parametric functional estimation and mixture models.

Copyright © 2008 Inderscience Enterprises Ltd.

1 Introduction

Supplier-manufacturer relationship has been a topic in academic research and manufacturing practice for a long time. The growing importance of Supply Chain Management (SCM) for a manufacturing company in a competitive environment entices researchers to examine this topic with a broader view. In this study, we analyse an important activity related to a manufacturer's Just-In-Time Purchasing (JITP) strategy under the digital economy. We model impacts of information sharing across a three-level chain (supplier-manufacturer-buyer) to a manufacturer in vendor evaluation[1] under the JITP practice. We measure scrap rates (during the manufacturing process) of a product's major components from different suppliers for quality performance, which is one of the most important criteria (other criteria include price, delivery, service, etc.) of vendor evaluation. We hope that the results of this study can provide practitioners some useful insights in transferring the value of information to quality improvement, cost reduction and customer satisfaction.

An organisation's purchasing activity is defined as the decision-making process by which formal organisations establish the need for products, identify, evaluate and choose among alternative brands and suppliers (Webster and Wind, 1972). JITP is the transfer of Japanese Just-In-Time (JIT) production concept to purchasing. For quality, JITP encourages interexchange between suppliers' and manufacturers' quality assurance people. Vendor evaluation based quality performance has a pivotal position in the JITP practice.

Quality is a primary factor to be considered by many buyers when shopping for products and services. For manufacturers, achieving economy, efficiency, productivity and quality concurrently in production is an important objective for business success. It is commonly understood that the goals in quality management are positively correlated to performance, as an example good quality usually leads to financial benefits (Montgomery, 1991). In quality related studies, two basic concepts are frequently used: defect rate and yield rate. When defective parts (or non-conforming parts as used in some quality literature) cannot be reworked into acceptable products, they are just scrapped. The defect rate is thus defined as the percentage of scrapped items in total products represented by 'δ' and the yield rate 'ξ', the counterpart of 'δ' (i.e. $\xi = 1 - \delta$). In this study, scrap rate in lieu of defect rate, which is more specific for practicing managers, is used throughout this paper.

From a manufacturer's perspective, quality related costs (including prevention costs, appraisal costs, internal failure costs and external failure costs) have direct impact to its financial performance. Costs that are caused by scraps and reworks belong to the category of internal failure costs (Montgomery, 1991). Oftentimes, those costs are the outcome of defective components from suppliers and can be significant (up to 40% of the total production costs). One of the motivations of this study came from a consulting project done in a pump manufacturing company by the first author. Customised pump products in the company have unique cost structures as the cost of the pump body in a pump is usually over 50% of the total production cost for the pump. Thus, replacing a scraped pump body means increasing production cost by at least 50%. For high value single pump projects (e.g. for those used in large hydro-power generator stations), this can easily turn project profits into losses.

When considering Quality Control (QC) process as a system (including inputs, processes and outputs), Montgomery (1991) categorised inputs into those of controllable and uncontrollable. He also pointed out that quality of raw materials and components from external suppliers is an important uncontrollable input in a manufacturing process (e.g. controllable inputs include self-manufactured components, QC process, skilled workers, etc.). However, there is no clear cut difference between controllable and uncontrollable inputs. For instance, information as a factor falls into both groups.

Information is the bridge between uncertainty and certainty. Sufficient information leads to efficiency and effectiveness, while distorted information can make things worse off. This was first studied by Forrester (1961) in his breakthrough 'industrial dynamics' and recalled with the emerging SCM research in the early 1990s as the 'bullwhip effect' (Lee et al., 1997a,b). Recent development in Information Technology (IT), especially technical infrastructures such as Enterprise Resource Planning (ERP) systems, has greatly facilitated the process of information storing and retrieving, which in turn stimulates the research interests in information sharing. For example, without Electronic Data Interface (EDI) and ERP, it would be almost impossible for a company to implement new approaches in SCM, like Cooperated Planning, Forecast and Replenishment (CPFR). For more detailed description of these innovations, readers are referred to recent research collections in SCM (Chen, 2003).

Generally speaking, there are three ways to classify research on information sharing. Firstly, most topics in information sharing can be classified based on their objectives in the study: how to avoid distorted information or how to quantify the value of shared information. Also, the close relationship between information and decision-making further leads to the classification according to types of control for systems (e.g. inventory): centralised versus decentralised control. Finally, coordination mechanisms differ under three situations of information availability: no information, partial information and full information (Sahin and Robinson, 2002).

Information is valuable for decision-making under the supply chain environment. However, realising the value of information is not an easy job. In the literature, we are aware of three approaches utilising shared information. This can be illustrated through studies in the fashion goods industry. The first approach applies multivariate demand distributions to explore the correlation of demand between the current and the future period (Fisher and Raman, 1996). The second approach uses Bayesian updating and stochastic dynamic programming (Eppen and Iyer, 1997). And the third approach is Bayesian Markov Decision Process (MDP), which is similar to the second one, but more elaborated (Ding et al., 2002).

Many authors have explored the topic of vendor evaluation in the literature. Some provided general frameworks for important elements in vendor evaluation, such as price, quality, delivery and service (Weber et al., 1991). Others analysed specific elements in more details (Chaudhry et al., 1993). They have greatly enhanced our understanding of purchasing, an important activity in SCM. However, so far, very little attention has been paid to the relationship between information sharing and vendor evaluation. This has motivated the present work and we intend to investigate this old topic under a new setup with a focus on quantitative analysis and managerial interpretation.

2 Modelling information sharing for JITP vendor evaluation

Under the diverse global business environment and market competition, manufacturers have to maintain relationships with competent suppliers, replace incompetent suppliers and develop new suppliers. Information is the key to those important decision-makings. However, oftentimes it is infeasible waiting for information to make decisions in an efficient and effective way. In this section, we follow the traditional research in this area to model three different scenarios of information sharing: no information sharing, partial information sharing and complete information sharing. However, definitions of these scenarios and approaches for related solutions in this study are slightly different from those in the existing literature.

Based on the first author's long-term industry consulting experience, it is observed that one important difference between QC under the traditional manufacturing environment and that of SCM is the changing focus of product life cycle. Under the traditional manufacturing environment, many manufacturers limit monitoring the quality of components from the suppliers during the manufacturing process within the enterprise 'four walls'. Beyond this scope, it can be the problem of rebates from the suppliers or warranties to the customers. Therefore, Statistical Process Control (SPC) is based on the assumption that quality information about the components obtained within the manufacturing company is 'complete' and process improvements depend on statistical analysis of those 'complete' data.

For SCM, market competition forces manufacturers to trace any opportunities in quality improvement, even beyond the enterprise 'four walls'. The focus has thus shifted from the manufacturer's quality management within the company to a group's (the manufacturer, its suppliers and customers) cooperated efforts in improving the quality of components across the whole supply chain. The quality related information from both its suppliers and customers can thus help the manufacturer better understand the quality problems related to the components. At the same time, similar information from the manufacturer can also help its suppliers to improve quality of the components and its customers to improve maintenance of the products related to those components. This usually leads to a win-win strategy (Chen, 2003).

We apply an approach that integrates both survival analysis, as widely used in Biostatistics (Kalbfleisch and Prentice, 1980) and reliability study (Meeker and Escobar, 1998) to model three different scenarios of information sharing. To the best of our knowledge, this is the first attempt of applying this approach to modelling JITP related activities. Below, we describe data and models in more details.

We begin with following notations to define the modelling process:

n	Number of suppliers
m_i	Number of batches of components from supplier 'i'
l	Number of stages
i	Index for suppliers (subscript)
$j(i)$	Index for batches of components from supplier 'i' (subscript)
t	Index for stages (superscript, use 'S' for suppliers' site, and 'C' customers' site)
$X^t_{ij(i)}$	Number of Scrapped components in batch '$j(i)$' from supplier 'i' at stage 't'
$Y_{ij(i)}$	Size of batch '$j(i)$' from supplier 'i'

$R_{ij(i)}^{t}$ Scrap rate of batch ‘*j(i)*’ from supplier ‘*i*’ at stage ‘*t*’

λ_{i}^{t} Mean of Poisson distribution for number of scrapped components from supplier ‘*i*’ at stage ‘*t*’

Notice that $X_{ij(i)}^{t}$ is used for historical quality data in most data sets and $R_{ij(i)}^{t}$ is used for scrap rate prediction in the vendor evaluation. The relationship between $X_{ij(i)}^{t}$ and $R_{ij(i)}^{t}$ is as follows:

$$R_{ij(i)}^{t} = \frac{X_{ij(i)}^{t}}{Y_{ij(i)}} \tag{1}$$

2.1 No information sharing

When a manufacturer knows nothing other than the quality history of components within the company, but recognises the fact that there should be defective components at the supplier’s manufacturing process as well as at the customer’s operation process within certain period (say, warranty period), then it has the so-called ‘truncated’ quality data and it is the typical scenario of ‘no information sharing’.

Under this situation, manufacturer can only use those internal quality data obtained within the company (not those from its suppliers or customers) to rank vendors according to the scrap rates at each manufacturing stage. However, unless other explanatory variables become available, the data can only be used by the manufacturer to predict the quality of components from potential vendors through the worst case analysis approach. For the pump manufacturer in our study, customer orders for standard pumps vary from tens to thousands. Wholesalers and retailers usually order in small to medium sizes when replenishing their inventories. On the other hand, industry customers with big projects may order in large sizes with a tight delivery lead time requirement. Consequently, the pump manufacturer may choose single JITP supplier for orders with small quantities and multiple JITP suppliers for those with large quantities. When there is an idle capacity in its own casting forge, the pump manufacturer may also produce the pump bodies by itself.

Studies in the QC department of the company with historical pump body scraping data illustrate that as a kind of rare event, most scrapings occur during the first and second stage in the manufacturing process, with a significant decrease in later stages. Thus, the number of defective components during the manufacturing process within the company can be modelled as a Poisson process. Furthermore, the number of defective components at each manufacturing stage can be modelled as a Poisson Random Variable (RV). However, as there is a proportional relationship between the time and the number of defective components (i.e. at the later manufacturing stage, the scrap rate on average becomes smaller), the rate of the Poisson process is also time-dependent. In statistical terminology, it is a non-homogeneous Poisson process.

For each stage of the manufacturing process, the number of defective components in each batch from each supplier is a RV under a Poisson distribution:

$$X_{ij(i)}^t \sim \text{Poisson}\left(\lambda_i^t\right) \tag{2}$$

It is well known that for a Poisson distribution, the parameter ('λ') equals to its mean and its variance. Thus, we can apply the Maximum Likelihood Estimation (MLE) approach to estimate 'λ' in a Poisson distribution and obtain an Upper Bound Value (UBV) with a certain confidence level (e.g. s = 0.95) for the number of defective components (or scraps):

$$\hat{\lambda}_i^t = \bar{X}_{ij(i)}^t = \frac{\sum_{j(i)=1}^{m_i} X_{ij(i)}^t}{m_i} \tag{3}$$

$$\mu\left(\bar{X}_{ij(i)}^t\right) = \hat{\lambda}_i^t + Z(s) \times \sqrt{\hat{\lambda}_i^t} \tag{4}$$

Note that in (4), $Z(s)$ is the Z value of the standard Normal distribution that generates a cumulative probability equal to the confidence level (e.g. a Z value of '1.64' related to a confidence level of '95%'). Then, using the relationship between scrap rate (R) and number of scraps (X), we can predict the UBV of the average scrap rate (R) for each supplier at each stage as follows:

$$\bar{R}_{ij(i)}^t = \frac{\mu\left(\bar{X}_{ij(i)}^t\right) \times m_i}{\sum_{j(i)=1}^{m_i} Y_{ij(i)}} \tag{5}$$

As JITP promotes both long-term relationship with selected suppliers and frequent delivery of small sizes, it is convenient to apply those Equations (3)–(5) ranking existing suppliers based on the predicted scrap rates at each stage. For the pump manufacturer in this study, the scrap rates at later manufacturing stages are more important based on the manufacturing cost structure and the lead time constraints for delivery.

2.2 Partial information sharing

When a manufacturer plans to expand its JITP supplier pool, it usually chooses to collect related information for the purpose of initial screening. The information (used as explanatory variables) may include, for example, production history, facility condition, quality certification and process policy. Together with corresponding information about its existing suppliers, manufacturer thus owns what we have previously described, 'censored' quality data. For most non-private companies, those related information can be obtained from public channels without direct communication. However, at this stage, external quality information (e.g. scarp rates at the suppliers' site) of components or products is still unavailable from either the existing suppliers or the existing customers. Hence, we can define this as the partial information sharing scenario.

We may apply Proportional Hazard (PH) regression model (Kalbfleisch and Prentice, 1980) with information from additional explanatory variables to calculate the average of scrap rates for a group of potential suppliers. In order to apply this model, we first discuss some terms used in modelling 'censored' data, that is, 'right censored', 'left censored' and 'interval censored'.

For the pump manufacturer who has all relevant information about its suppliers and customers, as well as varied direct quality information about the pump bodies, there are three possibilities. In case that the pump manufacturer has no information about external failures of the pump bodies, it owns so-called 'interval censored' data. The difference between 'interval censored' data and another commonly used term, 'truncated' data is the additional information about the explanatory variables. If the pump manufacturer has only information about failures of the pump bodies from its suppliers, it owns 'right censored' data. Similarly, if the pump manufacturer has only information about failures of the pump bodies from its customers, then it owns 'left censored' data.

The statistical methods available for analysing censored data vary according to types of censoring involved. In this study, we apply the PH regression model (Cox, 1972) to treat the 'interval censored' data (PH regression model is originally derived for 'right censored' data). For computational efficiency consideration, we can use PH regression model for 'interval censored' data. In practice, for two reasons we can approximate the 'interval censored' data as the 'right censored' data:

1 chances of scraping at the customers' side during the warranty period are small and

2 focus of this study is on the supplier-manufacturer relationship.

Hence the scraping at the suppliers' side is more important for vendor evaluation. We consider the failure time of pump bodies at each stage. Note that it follows a discrete survival-time distribution. This is feasible as some defects are deep inside the pump bodies and can only be found out during the process at later stages.

Define 'T' as the survival time taking values '$x_1 < x_2 <\ldots$' with probability density function *f(t)* and the cumulative distribution function *F(t)*. Then we have the expression of hazard λ_i^2 at x_i as follows:

$$\lambda_i = P\left(T = x_i \mid T \geq x_i\right) = \frac{f\left(x_i\right)}{1 - F\left(x_i\right)}, \quad i = 1, 2, \cdots \tag{6}$$

Under the initial condition '$F(0) = 1$', it is easy to see that this Ordinary Differential Equation (ODE) has solutions for $F(t)$ and $f(x_i)$ as follows:

$$F(t) = \prod_{i|x_i<t}\left(1 - \lambda_i\right) \tag{7}$$

$$f\left(x_i\right) = \lambda_i \prod_{j=1}^{i-1}\left(1 - \lambda_j\right) \tag{8}$$

We use the famous *Kaplan-Meier estimator* to estimate the survivor function (7). Suppose that d_i items fail at $x_i (i = 1, 2,\ldots,k)$ and m_i items are censored in the interval $[x_i, x_{i+1})$ at time spots $x_{i_1}, x_{i_2}, \ldots, x_{i_{mi}} (i = 0,1,\ldots,k)$, where $t_0 = 0$ and $t_{k+1} = \infty$. Let '$n_i = (m_i + d_i) + \cdots + (m_k + d_k)$' be the number of items at risk at a time just prior to t_i, then the *Kaplan-Meier estimator* of $F(t)$ is defined as (Kalbfleisch and Prentice, 1980):

$$\hat{F}(t) = \prod_{i|t_i < t} \left(\frac{n_i - d_i}{n_i} \right) \tag{9}$$

So, the Kaplan-Meier survivor function is a decreasing step function in the interval (0, 1).

While the general PH regression model is $\lambda(t; z) = \lambda_0(t) \exp(z\beta)$ (z is a row vector of $\boldsymbol{s}$ measured covariates, $\boldsymbol{\beta}$ is a column vector of s regression parameters and '$\lambda_0(t)$' is an arbitrary and unspecified base-line hazard function), the discrete version of the conditional survival function is

$$F(t;z) = \left[F_0(t)\right]^{\exp(z\beta)} \tag{10}$$

where $F_0(t)$ represents the base-line survival function for '$z = 0$'.

When there are ties, a linear log odds regression model can be used instead of the 'interval censored' data model (Kalbfleisch and Prentice, 1980):

$$\frac{\lambda(t;z)\mathrm{d}t}{1-\lambda(t;z)\mathrm{d}t} = \frac{\lambda_d(t)\mathrm{d}t}{1-\lambda_d(t)\mathrm{d}t}\exp(z\beta) \tag{11}$$

where $\lambda(t;z)\mathrm{d}t = 1-[1-\lambda_d(t)\mathrm{d}t]^{\exp(z\beta)}$.

From the direct scrap rate estimation to the PH regression model, additional information on explanatory variables can be used for the purpose of 'cause-and-effect' data analysis. Statistically, this can lead to increased accuracy in scrap rate prediction for new suppliers. For the pump manufacturer in this study, a comprehensive evaluation is processed for any JITP supplier to be a qualified supplier with quality certifications. Hence, the additional information on explanatory variables is available in both its QC department's documentation and its ERP system's '*Vendor*' database.

2.3 Complete information sharing

For a manufacturer under the SCM environment, when the IT infrastructure is capable to integrate the operational data across the supply chain, it may have the scenario of complete information sharing. For example, when a quality management module in ERP systems is used by both the manufacturer and its JITP suppliers, it is possible for the manufacturer to retrieve quality related data for the components directly from its suppliers' ERP systems. Also, sales and distribution module in ERP systems used by the manufacturer can detect the quality related information for the components from its customers' ERP systems through warranty service.

The advantage of complete information can be achieved by using the log-linear regression model used for count data analysis as described in Cameron and Trivedi (1997). The log-linear model uses a pseudo linear predictor '$\eta = Z^T\beta$' with the link function '$\eta = 1\mathrm{n}(\mu)$', which relates the linear predictor 'η' to the mean of the distribution 'μ'. Using the Generalised Linear Model (GLM) approach introduced by McCullagh and Nelder (1983), we can use the model to predict any specific scrap rate of pump bodies for each potential supplier through linking the random component with systematic components for independent observations, which follow a probability distribution (e.g. Normal, Beta).

For the pump manufacturer in this study, predicted values from the regression model are average scrap rates of pump bodies from a specific supplier. There are also internal and external data available to predict the scrap rates at each stage of the product life cycle across the supply chain. However, for the manufacturer, usually only predicted scrap rates within the company make sense for its vendor evaluation process.

For the convenience of data analysis, we use '$R_{ij(i)}^{t}, t = S, 1, 2, \ldots, l, C$'to represent the scrap rates for batch '***j***' from supplier '***i***', where '***S***' and '***C***' are respectively the manufacturing stages at suppliers' side and the warranty period at customers' side. We denote the mean value of the scrap rate distribution for supplier '***i***' at stage '***t***' as

$$\mu_i^t = E_{j(i)}\left(R_{ij(i)}^t\right), \quad i = 1, 2, \ldots, n; \quad t = S, 1, 2, \ldots, l, C \tag{12}$$

The log-linear model implies that

$$\ln\left(\mu_i^t\right) = \eta_i^t = \left(Z_i^t\right)^T \beta, \quad i = 1, 2, \ldots, n; \quad t = S, 1, 2, \ldots, l, C \tag{13}$$

where Z_i^t is the vector of independent coefficients representing the explanatory variables and $\boldsymbol{\beta}$ is the covariate parameter. It is well known in statistics that the estimates of μ_i^t can be obtained using the MLE approach. Once we know $\boldsymbol{\beta}$, we can predict μ_i^t through Z_i^t using Equation (13).

3 An example in ATO manufacturing

In order to show how the approach described in the previous section works in practice, we investigate a three-level supply chain (supplier-manufacturer-buyer) with a manufacturer under an Assemble-To-Order (ATO) manufacturing environment. We first describe the salient characteristics of the ATO manufacturing based on Song and Zipkin (2003). Then we discuss information sharing across the supply chain for components' quality used in ATO manufacturing by the pump manufacturer.

An ATO system includes several components and several products. Components are used to assemble different products. A special case is Engineering-To-Order (ETO), where components are partitioned into subsets and the customer selects components from those subsets (Song and Zipkin, 2003). In a pump manufacturing company, manufacturing of small standard pumps belongs to ATO are standardised by the manufacturer. However, when manufacturing of big customised pumps belongs to ETO, the engineering department needs to design the components as well as the configuration of components according to the customer's specific requirements. Obviously, an important difference between ATO and ETO is the lead time. The ETO system has a longer manufacturing lead time as it involves design lead time.

Manufacturing processes in the ATO system vary. The simplest situation is a single-period manufacturing process, like a pure assembly or distribution system. A more complex situation includes a multiple-period discrete manufacturing process, like the five-stage pump manufacturing process in this study. The most complex situation is possibly a continuous manufacturing process, like those in the chemical industry. When certain decision variables have stochastic properties, modelling those processes becomes substantially difficult (Song and Zipkin, 2003). The commonly used approach is modelling the process as a queueing system and the service measure of interests is the fill rate with time window 'τ', the probability of filling a demand within time 'τ', where 'τ' is a given non-negative integer. However, in this study, the measure we used is the scrap rate 'δ' (or the yield rate 'ξ').

The results presented here are based on an operation at a single facility of a pump manufacturer with additional self-manufacturing capacity (a casting forge) for major components – pump bodies. This facility is owned by a company belonging to a large international pump manufacturing group (annual sales around US$350 million). The company produces various types of pumps for both domestic and overseas customers. There are six series of standard centrifugal pump products, as well as other customised pump products. The company uses the JITP strategy to do business with its suppliers for pump bodies.

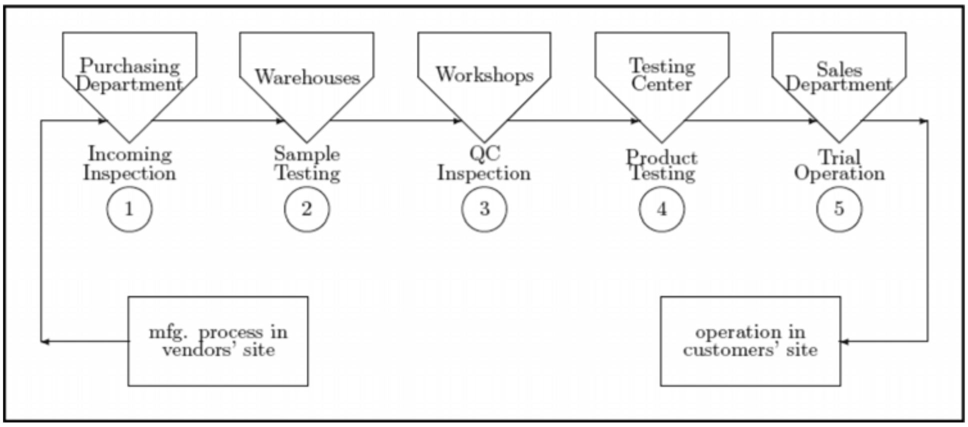
A typical five-stage (we define stages here according to activities related to the QC process for pump bodies) ATO system is used for standard pump products. The process works as follows:

1. Raw materials and components from suppliers are inspected by the purchasing department before moving to various inventory locations.
2. Sample testings for the pump bodies are operated by the QC department before moving non-defective pump bodies to different workshops.
3. QC inspections are operated after each stage in the manufacturing process in the workshops, including machining, painting, assembling and labelling.
4. Product testings are operated in the pump testing centre before moving finished pumps to product warehouses.
5. Finished pumps are delivered to various customers' locations by the sales department for installation and trial operation according to the sales contracts.

At each stage, defective pump bodies are identified and scraped, as no rework can be done for Costing material. Furthermore, beyond the scope of the manufacturing process, there are two interfaces with external controls in a supply chain: the suppliers and the customers. A simple flow chart of the steps involved in the processing operation at this pump manufacturer is shown in Figure 1.

Due to variety and volume of its pump products, the pump manufacturer has developed a group of local suppliers for its JITP practice. It also orders pump bodies from its suppliers as a local agent for its foreign group members from time to time. To satisfy increasing business needs, the pump manufacturer is continuously accepting new suppliers into its JITP supplier pool. As a result, one important task of its purchasing department is just the vendor evaluation.

Figure 1 Pump manufacturer's five-stage manufacturing process



A weighted evaluation system which includes measures about price, quality, delivery and service is used. This weighted evaluation system uses an index calculated by weights and measures under four categories with data from the transaction system under ERP. Each year, QC manager sets weights for price (0.2–0.3), quality (0.4–0.5), delivery (0.1–0.2) and service (0.1–0.2) after consulting with other functional departments. Obviously, quality of those pump bodies is the major concern for two reasons:

1 the chemical property of casting materials that no rework can be done and

2 the manufacturing lead time to reproduce those casting materials that are significant for the manufacturing process.

The measures used here are the averages of scrap rates (for quality), the ratio between actual purchasing price and industry reference price (for price), the percentage of delay in delivery (for delivery) and the response rate for service requests (for service).

Information sharing across the supply chain for the pump manufacturer varies according to the IT infrastructure for the supply chain. While the manufacturer has been using EDI system for a long time and has recently implemented the ERP system as well, information sharing across the supply chain is still limited. However, the pump manufacturer is interested in improving its productivity through information sharing across the supply chain. As a result, it begins to collect and utilise quality related information for pump bodies within the manufacturing process as well from its suppliers and customers in the ERP systems for further analysis.

More specifically, the information sharing with its suppliers is accomplished through regular plant visits as defined in the supplier development programme for JITP. On the other hand, the information sharing with its customers is achieved through its customer service programmes. For example, its sales department can collect quality related information during the on-site installation and trial operation period. Also, as a condition of the warranty, the customers are required to provide detailed information for any quality problem that leads to after-sales customer service within the warranty period.

To show how we can use those models in practice, we developed algorithms for three scenarios based on the stochastic models as discussed in the previous section.

And to illustrate how to use those models for the pump manufacturer's decision making through algorithms implemented using a computer system, we used a small sample data set from database in the ERP system for a numerical study. Typical results from the numerical study work as a snapshot for the actual system. Furthermore, we provided step-by-step simulation solutions for different cases and detailed descriptions about related routines in statistical software (see Appendix). Those programs will be helpful for practicing managers when designing similar systems.

Firstly, we have the number of scraps in pump body batches supplied by the existing suppliers, namely supplier '*A*', '*B*', '*C*', '*D*' and '*F*' (see Table 1). For the purpose of visibility, number of pump bodies in each batch has been rounded to the nearest multiples of ten units. Table 1 also contains information for those suppliers under the complete information sharing scenario. To simplify the illustration, we will reuse related information for statistical analysis under other scenarios.

Table 1 Number of pump body scraps for existing suppliers

Supplier	*Batch*	*Quantity*	*Number of Scraps*						
			S	*I*	*II*	*III*	*IV*	*V*	*C*
A	1	40	0	3	1	0	2	1	1
	2	60	1	5	2	1	1	2	0
	3	100	2	8	3	4	3	1	2
	4	50	1	2	0	1	1	0	0
	5	100	1	7	2	0	2	1	0
	6	40	0	4	2	1	0	0	1
B	1	15	0	1	0	2	1	0	1
	2	20	1	1	1	0	1	1	0
	3	10	0	0	1	0	2	0	0
	4	10	1	1	2	1	0	2	1
C	1	10	0	1	0	1	0	0	1
	2	10	0	0	2	0	0	1	0
	3	10	1	0	0	1	0	0	0
D	1	40	2	1	4	0	2	1	1
	2	20	1	0	1	1	0	2	0
	3	10	0	1	2	0	1	0	0
	4	20	1	1	0	2	1	0	1
F	1	150	2	8	3	4	1	0	3
	2	100	1	6	2	1	0	2	2
	3	100	1	5	4	2	1	1	1
	4	50	0	3	2	3	0	0	1

Note: Where 'S' represents the manufacturing stage at the supplier's location and 'C' represents the warranty period at the customer's location.

Secondly, we also have some quality related information for both the existing and new suppliers (see Table 2).

Table 2 Quality related information about existing and new suppliers

Information	*Supplier*							
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Relation (R)	*P*	*R*	*P*	*S*	*P*	*P*	*R*	*S*
History (H)	15	4	20	10	4	2	10	10
Output (O)	1500	500	300	100	400	50	1000	50
TQM Status (T)	*Y*	*N*	*Y*	*Y*	*N*	*Y*	*Y*	*N*

Note: Where the meaning of the information is as follows:

1. ‘Relation’ is the relationship with the manufacturer, including: *S* as Special with value ‘0’; *P* as Partner with value ‘1’ and *R* as Regular with value ‘2’.
2. ‘History’ is the history of the production for the components by supplier, in years.
3. ‘Output’ is the yearly maximum production capacity for the components, in units.
4. ‘TQM Status’ is the status of the TQM certification, including: *Y* as yes with value ‘0’; *N* as No with value ‘1’.

We apply the ‘worst case analysis’ approach for UBVs of the pump body scrap rates in our analysis. These UBVs are generated through simulation-based prediction method for the censored data as introduced in Escobar and Meeker (1999). This method is feasible to be implemented in computing and also satisfies the pump manufacturer’s actual requirements. As the pump manufacturer is interested in predicting the bottom line of quality for new suppliers based on data from the current suppliers, the UBV becomes a suitable measure. It also makes sense to compare different suppliers by using this measure.

The operational data related to this study exist in several master data files in the pump manufacturer’s ERP system. To run these algorithms, the first step is to retrieve required data and to consolidate them into a customised dataset for the purpose of statistical analysis. This is achieved by the IT infrastructure. Under the operation with ERP system, the purchasing department is responsible for entering purchase orders and receipts for the pump bodies, the QC department maintains the scraping records during the manufacturing process and the sales department records requests for warranty from the customers into a centralised database system. Thus, the pump manufacturer has related information about the suppliers and customers, number of batches supplied, number of pump bodies in each batch and number of scraps at each stage throughout the whole manufacturing process.

The computational schemes (can be implemented in many statistical packages, use S-Plus for illustration here, see Appendix for more details) for three scenarios using simulation-based prediction methods are as follows.

No information sharing scenario:

1 Use Equation (3) to compute the average of scraps at each stage for batches of pump body from each supplier. Those values represent estimated mean values ('λ') of related Poisson distributions.

2 Use Equation (4) to generate UBVs of the average scraps through simulation.

3 Use Equation (5) to compute related scrap rates.

Partial information sharing scenario:

1 Determine survival time of pump bodies based on number of scraps at each stage of the manufacturing process. For example, if the number of scraps in stage three is two, it means that two pump bodies have survival time of three. This way, we will have the survival time for each individual pump body.

2 Use S-Plus routine '*coxph*' for the PH regression based on the survival time data from step 1 and related information about explanatory variables.

3 Identify important explanatory variables in the regression model based on the '*p*-value' in the summary table generated by the '*coxph*' routine.

4 Use S-Plus routine '*survfit*' to obtain the predicted average yield rate 'δ' and related Lower Bound Value (LBV).

5 Apply the relationship between the yield rate 'δ' and scrap rate 'ξ' ($\xi = 1-\delta$) to obtain UBV of the scrap rates. This is the UBV of predicted scrap rate for any new supplier.

Complete information sharing scenario:

1 Same as Step 1 under the partial information sharing scenario.

2 Use S-Plus routine '*glm*' to run the regression based on survival time data from step 1 and related information about the explanatory variables.

3 Identify important explanatory variables in the regression model based on the '*p*-value' in the summary table generated by the '*glm*' routine.

4 Based on the revised regression model as in step 3, predict the average of survival time for pump bodies from individual new supplier using information about explanatory variables.

5 Apply '*Little's Law*'[3] to obtain the average of scrap rates for pump bodies from each new supplier. Those values give us estimated mean values of related Poisson distributions.

6 Use Equation (4) to generate UBVs of average scrap rates for each individual new supplier.

After running algorithms for each scenario, we have the following results:

1 *No information sharing*: based on the result under an equal weighted system (see Table 3), we can directly rank the existing suppliers under a priority consequence (notice the fact that the smaller the weighted scrap rate, the better the quality) as: '***F***', '***A***', '***D***', '***B***' and '***C***'. In the practical system, ranking has

various weighting options. For example, if the pump manufacturer believes that scraping at later stages is more significant, then the weights can be chosen as ‘0.1’(Stage 1), ‘0.1’(Stage 2), ‘0.2’(Stage 3), ‘0.3’(Stage 4) and ‘0.3’(Stage 5), respectively.

2 *Partial information sharing*: notice that in this case the scrap rate at the second stage is most significant (see Table 4). This can be the bottom line for any new supplier under the ‘worst case analysis’. It becomes useful when the pump manufacturer needs to order additional pump bodies from random new suppliers in rush under a situation when all existing JITP suppliers have run out of their production capacity limits.

3 *Complete information sharing*: notice that we only selected four categories here. Thus the regression result will be fairly close (see Table 5). However, the pump manufacturer can still rank new suppliers according to minor differences. Under the priority consequence with an equal weighted system, it can choose consequently (notice the fact that the smaller the weighted scrap rate, the better the quality): Supplier ‘***H***’, ‘***G***’ and ‘***E***’.

Table 3 Predicted 95% UBVs of pump body scrap rates: existing supplier

Stage	*Weight*	*Supplier*				
		A	*B*	*C*	*D*	*F*
I	0.2	0.0625	0.1311	0.1782	0.0836	0.0339
II	0.2	0.0296	0.1868	0.1049	0.1053	0.0234
III	0.2	0.0260	0.1142	0.0949	0.1514	0.0449
IV	0.2	0.0259	0.1548	0.1163	0.0705	0.0109
V	0.2	0.0313	0.2034	0.1025	0.0976	0.0174
Weighted value		0.0351	0.1581	0.1194	0.1016	0.0261
Rank		2	5	4	3	1

Table 4 Predicted 95% UBVs of pump body scrap rate: new suppliers’ group

Supplier	*Stage*				
	I	*II*	*III*	*IV*	*V*
New supplier’s group	0.0891	0.1053	0.0615	0.0257	0.0261

Table 5 Predicted averages of pump body scrap rate: individual new supplier

Stage	*Weight*	*Supplier*		
		F	*G*	*H*
I	0.2	0.1242	0.1229	0.1183
II	0.2	0.0295	0.0289	0.0271
III	0.2	0.0052	0.0051	0.0046
IV	0.2	0.0007	0.0007	0.0006
V	0.2	0.0001	0.0001	0.0001
Weighted value		0.0392	0.0351	0.0301
Rank		3	2	1

Overall, the results in the numerical study provided us a clear picture about the value of information sharing. Without information sharing, the pump manufacturer may only compare existing suppliers. Under partial information sharing, it can evaluate the worst case of any new supplier's quality for its pump bodies. Furthermore, under complete information sharing, it will be able to compare individual new suppliers. More important, this example illustrated the application of advanced analytical tools in aid of decision making under a data rich environment. Relevant enterprise data is used to increase the prediction accuracy under random distribution and regression models. This greatly enhances the value of ERP systems in the enterprise decision making process.

To test the robustness of above algorithms, we also tested different probability distributions and carried out a sensitivity analysis designed for Poisson distribution. The results from the simulation and sensitivity analysis provided us additional insights about limitations in the proposed approach and clues for future improvements. The models and computational algorithms have been validated in practice through customised commercial programmes. To protect the confidentiality of these programmes, the authors are not in a position not to release the technical details beyond the scope of this study.

The results (see Tables 6 and 7) of the simulation study and the sensitivity analysis illustrated the impact of different distributions to performance evaluation. Comparison with the real world example also provided us the justification in choosing non-homogeneous Poisson process as the tool for analytical system design. For small samples, the impact to the performance evaluation from changes in distribution parameters is not significant. More specifically, we have the following observations:

1 The assumption of different distributions has direct impact to the ranking result. There is an evident gap between the discrete distributions (i.e. Poisson and Binomial) and the continuous distribution (i.e. β).

2 When simplifying with time-homogeneous distribution and an equal weighted system, we only need to check the overall average of scrap rates for all stages.

3 For small samples, the regression result is relatively robust. However, important explanatory variables may change. Also, there is certain threshold (i.e. $\lambda = 0.04$) under certain distribution (i.e. Poisson) that will change the ranking result.

Table 6 Comparison study with different distributions

Distribution	*Average of scrap rates*					*Rank*				
	A	*B*	*C*	*D*	*F*	*A*	*B*	*C*	*D*	*F*
Part 1: no information sharing scenario										
Poisson	0.02	0.032	0.012	0.044	0.012	3	4	1/2	5	1/2
Binomial	0.012	0.028	0.012	0.036	0.02	1/2	4	1/2	5	3
Beta	0.036	0.02	0.012	0.02	0.028	5	2/3	1	2/3	4

Table 6 Comparison study with different distributions (continued)

Distribution	*Average of scrap rates*
	Bottom line for any new supplier
Part 2: partial information sharing scenario	
Poisson	0.0808
Binomial	0.0789
Beta	0.0723

Distribution	*Important variable*	*Average of scrap rates*			*Rank*		
		E	*G*	*H*	*E*	*G*	*H*
Part 3: complete information sharing scenario							
Poisson	rel, rqm	0.0982	–	0.2124	2	1	3
Binomial	rel, Inhis, Inout	0.112	0.112	0.112	tie	tie	tie
Beta	rel, Inout, tqm	0.01174	0.0534	0.0347	3	2	1

Note: Where 'rel' represents 'Relationship', 'Inhis' log value of 'History', 'Inout' log value of 'Output' and 'tqm' 'TQM Status'.

Table 7 Sensitivity analysis under Poisson distribution

Parameter	*Average of scrap rates*					*Rank*				
	A	*B*	*C*	*D*	*F*	*A*	*B*	*C*	*D*	*F*
Part 1: no information sharing scenario										
0.02	0.028	0.032	0.012	0.044	0.012	3	4	1/2	5	1/2
0.024	0.032	0.032	0.012	0.044	0.012	3/4	3/4	1/2	5	1/2
0.028	0.032	0.032	0.012	0.044	0.012	3/4	3/4	1/2	5	1/2
0.032	0.036	0.032	0.012	0.044	0.02	4	3	1/2	5	1/2
0.036	0.032	0.032	0.012	0.044	0.012	3/4	3/4	1/2	5	1/2
0.04	0.052	0.032	0.012	0.044	0.012	5	3	1/2	4	1/2

Parameter(λ)	*Average of scrap rates*
	Bottom line for any new supplier
Part 2: partial information sharing scenario	
0.02	0.0888
0.024	0.093
0.028	0.088
0.032	0.0867
0.036	0.0852
0.04	0.1062

Table 7 Sensitivity analysis under Poisson distribution (continued)

Parameter (λ)	*Important variable*	*Average of scrap rates*			*Rank*		
		E	*G*	*H*	*E*	*G*	*H*
Part 3: complete information sharing scenario							
0.02	rel, Inout, tqm	0.2426	0.0422	0.2874	2	1	3
0.024	rel, Inout, tqm	0.2539	0.0701	0.2865	2	1	3
0.028	rel, rqm	0.0198	–	0.1435	2	1	3
0.032	rel, rqm	0.0321	–	0.1541	2	1	3
0.036	rel, rqm	0.0483	–	0.1686	2	1	3
0.04	rel, Inhis, Inout	0.1499	0.0147	0.07	3	1	2

Note: Where 'rel' represents 'Relationsip', 'Inhis' log value of 'History', 'Inout' log value of 'Output' and 'tqm' 'TQM Status'.

The pump manufacturer's operation system in practice is far more complicated than what we have illustrated here. As quality is a part of the measuring system used in the vendor evaluation, ranking results from those algorithms further work as inputs to the customised vendor evaluation function in SAP R/3 system's Material Management (MM) module and to generate the overall ranking together with other important measures like price, delivery and service. Also, the explanatory variables in the regression model will be as many as tens to hundreds and the most recent results can be shown online whenever needed through Business Intelligence (BI) add-ins.

4 Conclusion

In this paper, we explored an analytical solution for information sharing under the JITP practice for an ATO system. For this particular pump manufacturer, the integration of ERP system and analytical tools developed using the stochastic models provided it opportunities to realise the value of information sharing in quality improvement, cost reduction and customer satisfaction (through its IT investment, i.e. the ERP system). This solution is unique for two reasons: the chemical property of the pump bodies and the JITP practice under the ERP environment. Extensions of the solution to other industries and manufacturing processes will depend on particular information sharing requirements as well as the external operation environment. However, the general analytical approach of treating information sharing in quality management may give both academic researchers and practical managers useful insights in designing similar integration systems.

More importantly, this study demonstrated an example of leading industrial practices in applying analytical solutions for key operational decision making under an advanced IT infrastructure. While most academic research in this area so far has been focused on illustrating the value of information sharing for SCM, companies are in great need of practical algorithms and decision tools to realise such value. For innovations in procurement such as JITP, managing complexity under diversity is critical to promise any advantages. IT infrastructure, like the ERP system, is a necessity for many companies fighting for the complexity in operation. However, the lack of 'planning' functions in the package software has been a major criticism for the ERP solutions (Davenport, 2000). The integration of the analytical solutions into the ERP system represents a trend for both package software vendors and users to achieve effective and efficient decision making under the data-rich environment.

As an initial effort, this study is also at a position to be improved under both theoretical and practical considerations. Under the theoretical paradigm, attentions should be paid to the statistical justification of probability distributions selected to represent the stochastic property of quality measures. For the pump manufacturer in this study, Poisson distribution and normal approximation are used in the simulation study to calculate the number of pump body scraps. For other processes and components, other distributions, such as Weibull, Exponential or Lognormal distribution, may also become suitable candidates.

On the practical side, many issues in the implementation should be considered. For the pump manufacturer, as SAS has been used in the production planning and control area, integration of the analytical package with the ERP system becomes a natural choice. However, for companies with only ERP systems, the selection and the integration of analytical packages can be an important topic in planning the overall IT infrastructure. Many third-party software vendors also provide analytical solutions to the ERP systems. It is then important for the companies to make buying versus self-development decisions for those analytical solutions.

Acknowledgement

We would like to thank the reviewers whose detailed questions and suggestions considerably improved this paper. The authors also thank Dr. Timothy Urban for his constructive suggestions and comments.

References

Cameron, C.A. and Trivedi, P.K. (1997) *Regression Analysis of Count Data*, Cambridge, UK: Cambridge University Press.

Chaudhry, S.S., Forst, F.G. and Zydiak, J.L. (1993) 'Vendor selection with price breaks', *European Journal of Operational Research*, Vol. 70, No. 1, pp.52–66.

Chen, F. (2003) 'Information sharing and supply chain coordination', in T. de Kok and S. Graves (Eds). *Supply Chain Management*, *Vol. 11 of Handbooks in Operations Research and Management Sciences*, Chapter 7, North-Holland, Amsterdam, The Netherlands.

Cox, D.R. (1972) 'Regression models and life table (with discussions)', *Journal of Royal Statistical Society*, Vol. 34, pp.187–220.

Davenport, T.H. (2000) *Mission Critical: Realizing the Promise of Enterprise Systems*, Boston: Harvard Business School Press.

Ding, X., Puterman, M.L. and Bisi, A. (2002) 'The censored newsvendor and the optimal acquision of information', *Operations Research*, Vol. 50, No. 3, pp.517–527.

Eppen, G.D. and Iyer, A.V. (1997) 'Improved fashion buying with Bayesian updates', *Operations Research*, Vol. 45, No. 6, pp.805–819.

Escobar, L.A. and Meeker, W.Q. (1999) 'Statistical prediction based on censored life data', *Technometrics*, Vol. 41, No. 2, pp.113–124.

Fisher, M. and Raman, A. (1996) 'Reducing the cost of demand uncertainty through accurate response to early sales', *Operations Research*, Vol. 44, No. 1, pp.87–99.

Forrester, J. (1961) *Industrial Dynamics*, Cambridge, MA: MIT Press.

Kalbfleisch, J. and Prentice, R. (1980) *The Statistical Analysis of Failure Time Data*, New York: Willey.

Lee, H.L., Padhamanabhan, V. and Whang, S. (1997a) 'The bullwhip effect in supply chain', *Sloan Management Review*, Spring, pp.93–102.

Lee, H.L., Padhamanabhan, V. and Whang, S. (1997b) 'Information distortion in supply chain: the bullwhip effect', *Management Science*, Vol. 43, No. 4, pp.546–558.

McCullagh, P. and Nelder, J.A. (1983) *Generalized Linear Models*, Cambridge, UK: Cambridge University Press.

Meeker, W.Q. and Escobar, L.A. (1998) *Statistical Methods for Reliability Data*, New York: John Wiley & Sons Inc.

Montgomery, D.C. (1991) *Introduction to Statistical Quality Control*, 2nd edition, New York: John Wiley & Sons.

Sahin, F. and Robinson, P.E. (2002) 'Flow coordination and information sharing in supply chain: review, implications, and directions for future research', *Decision Sciences*, Vol. 33, No. 4, pp.505–536.

Song, J. and Zipkin, P. (2003) 'Supply chain operations: assemble-to-order and configure-to-order systems', in T. de Kok and S. Graves (Eds). *Supply Chain Management*, *Vol. 11 of Handbooks in Operations Research and Management Sciences*, Chapter 11, North-Holland, Amsterdam, The Netherlands.

Weber, C.A., Current, J.R. and Benton, W. (1991) 'Vendor selection criteria and methods', *European Journal of Operational Research*, Vol. 50, No. 1, pp.2–18.

Webster, F. and Wind, Y. (1972) *Organizational Buying Behavior*, Englewood Cliffs, NJ: Prentice-Hall.

Notes

[1]Another commonly used term is 'supplier selection'. Although there may be minor difference between them, we use those two terms iteratively in this paper. This also applies to the terms 'supplier' and 'vendor'.

[2]This is different from the parameter 'λ' in a Poisson distribution of the previous discussion.

[3]'Little's Law' in the queue theory indicates that 'number' of transactions is the product of arrival 'rate' of transactions and 'time' in system of transactions. In this case, the 'number' is the average failure time, 'rate' the scrap weight and 'time' the stage. For example, if we have 1, 4, 1 and 1 scrap at stage 1–4, with total 7 scraps in the batch, then the scrap weights for each stage are 1/7, 4/7, 1/7 and 1/7 respectively and the average failure time is '$1/7 + 4/7 \times 2 + 1/7 \times 3 + 1/7 \times 4 = 2/3$'.

Appendix

Background about Cox PH regression model

Cox PH regression model is widely used in survival analysis for hazard function estimation. It uses a non-parametric approach which is based on an arbitrary base-line hazard function. First proposed by Cox (1972), the *Cox PH regression model* uses time dependent covariates to predict the survivor function. There are various approaches in applying MLE for parameters (i.e. 'β') and survivor function (i.e. 'λ') estimation. Readers interested in technical details see Kalbfleisch and Prentice (1980).

A computational walk-through

We use the example with Poisson distribution in the simulation study to process a computational walk-through. The objective of this detailed walk-through is to let practicing managers with average knowledge about analytical solutions have some ideas about the technical implementation in computation. Major routines in statistical packages used in the implementation are described in the next subsection.

The starting point of the simulation study is to use a random number generator (can be found in most statistical packages, in this case, uniform [0.01, 0.04]) to generate average of scrap rates for each existing supplier (i.e. *A*, *B*, *C*, *D* and *F*). For Poisson distribution, this is just the 'λ' value.

However, in order to compare with other distributions in the simulation study, we relaxed the assumption of non-homogeneous Poisson process in the real world example. Also, for the convenience of comparison study, we assume that all suppliers have supplied one batch of 50 pump bodies. Then, we can use another random number generator (i.e. Poisson [50, λ]) to generate number of scraps at each stage for each supplier. Once we have the distribution for the number of scraps, we are ready to go through different scenarios:

1 *No information sharing scenario*: ranking existing suppliers. This is the simplest situation. As we have distributions for number of scraps at each manufacturing stage from each existing supplier, we can as usual, apply the equal-weighted approach to calculate the average of scrap rates for each supplier (in this case, [0.02, 0.032, 0.012, 0.044, 0.012] for supplier [*A*, *B*, *C*, *D*, *F*]). Then, we can rank them (in this case, [3, 4, 1/2, 5, 1/2] for supplier [*A*, *B*, *C*, *D*, *F*]).

2 *Partial information sharing scenario*: predicting bottom line for all new suppliers. Under this scenario, we first need to transfer distributions for number of scraps into distributions for survival time of individual components (in this case, survival time is the number of stages a component going through before failure, i.e. 1, 2, 3, 4, 5, 6, 7, inclusive). We also need to add a status indicator variable for censored items (in this case, all qualified components are censored, as we do not know their quality status after the manufacturing process. Once the data transformation is ready, we simply apply 'PROC PHREG' routine in SAS to obtain the mean and standard deviation of the baseline survival time (in this case, '6.564' and '0.0785', respectively). Based on the relationship between survival time and yield rate, as well as the between yield rate and scrap rate

(in this case, (Average of Yield Rate) = (Average of Survival Time) ÷ (Number of Total Stages or 7) and (Average of Scrap Rate) = 1− (Average of Yield Rate)). As we are interested in UBV of scrap rates, we first obtain UBV of the survival time (i.e. '6.6936') using '6.564 + 1.64 × 0.0785' ('1.64' is related to the '0.95' confidence level). Then, we have the UBV of scrap rates as '0.0808'. This is the bottom line scrap rate the management can expect for any new supplier.

3 *Complete information sharing scenario*: predicting for individual new suppliers. Further by applying the survival time data with additional information from both the suppliers and customers for '*PROC GLM*' routine in SAS, we can identify '*Relationship*' and '*TQM* Status' as two important explanatory variables through the forward step selection procedure. Then, we can predict the average of survival time for each new supplier with the regression equation (i.e. (Average Survival Time) = 6.4474 + 0.799 × ('Relationship') − 0.9342 × ('TQM Status'), '[6.3123, –, 5.5133]' for new supplier [*E*, *G*, *H*] in this case.). As the last step, we obtain the average of scrap rates as in the previous scenario (i.e. '[0.0982, − , 0.2124]' for new supplier [*E*, *G*, *H*]) and rank them (i.e. [2, 1, 3] for new supplier [*E*, *G*, *H*]). Notice two things here: first, '–' indicates the out-of-range value, for the survival time, it is equal to or larger than '7' and for the scrap rate it is equal to or less than '0'; second, to minimise impacts from large numerical value, we used a log transformation for values of '*History*' and '*Output*' before fitting for the model and running for the regression.

Related routines in statistical packages

Survival analysis routine survfit (All routines in S-Plus are case-sensitive) in S-Plus.

'surfit' is used to compute an estimate of a survival curve for censored data or the predicted survival function for the Cox (1972) PH regression model. The basic formula used is as follows:

$$\text{survfit(formula, data, options)}$$

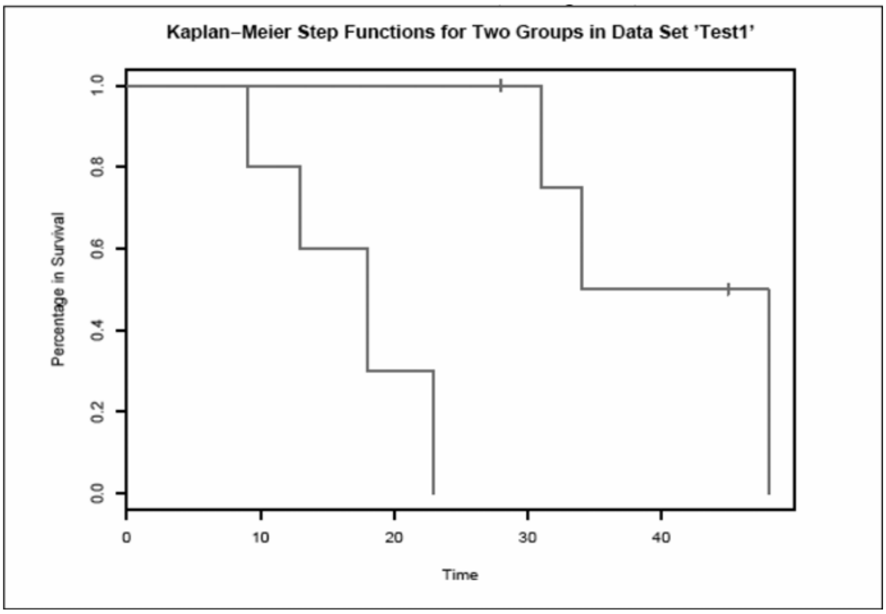
where 'formula' is a formula object or a 'coxph' object. For a formula object, a 'Surv' object should be included as the response on the left of the '~' operator and a group of independent variables on the right. 'data' is a data frame in which to interpret the variables named in the formula. 'options' decide information like 'weights', 'subset', 'type', etc.

Here is an example:

```
# Fit a Kaplan-Meier survivor function and plot it
test1 <- list(time = c(9,13,13,18,23,28,31,34,45,48), status =
c(1,1,0,1,1,0,1,1,0,1), x = c(1,1,1,1,1,2,2,2,2,2))
fit <- survfit(Surv(time, status) ~ x, data = test1)
plot(fit)
```

In this example, two survival curves based on a data set ‘test1’ for relationships between dependent variables ‘time’, ‘status’ and independent variable ‘x’ are plotted. It shows that for the group with ‘$x = 1$’, percentages of survival at time ‘9’, ‘13’, ‘18’ and ‘23’ are ‘0.8’, ‘0.6’, ‘0.3’ and ‘0’. Similar, for another group, percentages of survival at time ‘31’, ‘34’ and ‘48’ are ‘0.75’, ‘0.5’ and ‘0’ (see Figure A1).

Figure A1 Kaplan-Meier step functions for two groups in data set ‘test1’



A similar routine in SAS is ‘PROC LIFETEST’.

Survival analysis routine coxph in S-Plus

‘coxph’ is used to fit the Cox (1972) PH regression model. The basic formula is as follows:

$$\text{coxph(formula, data = parent.frame(), options)}$$

where ‘formula’ is a formula object with response in the ‘Surv’ function format on the left of a ‘~’ operator and explanatory variables on the right. ‘data’ is a data frame in which to interpret the variables in the formula. ‘options’ contain other information like ‘weights’, ‘subset’, ‘method’, etc.

Here is an example:

```
# Create a simple data set for a time-dependent model
test2 <- list(start = c(1,2,5,2,1), stop = c(2,3,6,7,8), event =
   c(1,1,1,1,0), x = c(1,0,0,1,0))
summary(coxph(Surv(start, stop, event) ~ x, test2))
```

In this example, the survivor function has an explanatory variable x and the last observation is censored at time ‘8’. The summary information about the Cox (1972) PH regression model for the data set ‘test2’ is provided by the routine. It shows that the coefficent for x is 0.347 and that for e^x is 1.41. Hence, the estimated survivor function value is given by ‘$y = 1.41\ e^x$’, (see Figure A2).

Figure A2 S-Plus ‘coxph’ routine output

```
Call:
coxph(formula = Surv(start, stop, event) ~ x, data = test2)

  n= 5
   coef exp(coef) se(coef)     z    p
x 0.347      1.41     1.02 0.341 0.73

  exp(coef) exp(-coef) lower .95 upper .95
x      1.41      0.707     0.193      10.3

Rsquare= 0.023   (max possible= 0.762 )
Likelihood ratio test= 0.12  on 1 df,   p=0.733
Wald test            = 0.12  on 1 df,   p=0.733
Score (logrank) test = 0.12  on 1 df,   p=0.732
```

A similar routine in SAS is ‘PROC PHREG’. Another SAS routine, ‘PROC LIFEREG’, can be used for left, right and interval censored data under the GLM approach.

Regression routine glm in S-Plus

‘glm’ is used to fit for a GLM model. The basic formula is as follows:

glm(formula, family = gaussian, options)

where ‘formula’ is a symbolic description of the model to be fit. The model includes a predictor under the form ‘response ~ terms’ where ‘response’ is the (numeric) response vector and ‘terms’ is a series of terms which specifies a linear predictor for ‘response’. ‘options’ contain other information like ‘data’, ‘weights’, ‘subset’, ‘method’, etc.

Here is an example:

```
# A Gamma example from McCullagh and Nelder (1983)
test3 < − data.frame(u = c(5,10,15,20), l1 = c(118,58,42,35), l2
= c(69,35,21,26))
summary(glm(l1 ∼ log(u), data = test3, family = Gamma))
summary(glm(l2 ∼ log(u), data = test3, family = Gamma))
```

In this example, summary information about two GLM models for the data set ‘test3’ is provided by the routine (see Figure A3).

$$l1 = -0.0140273 + 0.0139265[\log(\mu)]$$
$$l2 = -0.020064 + 0.021537[\log(\mu)]$$

Figure A3 S-Plus 'glm' routine output

```
Call:
glm(formula = l1 ~ log(u), family = Gamma, data = test3)

Deviance Residuals:
        1          2          3          4
-0.010419   0.045607  -0.005179  -0.031071

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0140273  0.0011797  -11.89  0.00700 **
log(u)       0.0139265  0.0006034   23.08  0.00187 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.001612063)

    Null deviance: 0.9284704  on 3  degrees of freedom
Residual deviance: 0.0031809  on 2  degrees of freedom
AIC: 21.053

Number of Fisher Scoring iterations: 3


Call:
glm(formula = l2 ~ log(u), family = Gamma, data = test3)

Deviance Residuals:
        1          2          3          4
 0.007281   0.033069  -0.211152   0.148396

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.020064   0.008817  -2.276   0.1507
log(u)       0.021537   0.004434   4.857   0.0399 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.0320441)

    Null deviance: 0.863737  on 3  degrees of freedom
Residual deviance: 0.067753  on 2  degrees of freedom
AIC: 29.234

Number of Fisher Scoring iterations: 4
```

A similar routine in SAS is 'PROC GLM'.