INFORMATION QUALITY COST CURVES

(Practice-oriented paper)

Anton Raneses
Oklahoma State University
anthony.raneses@okstate.edu

Michael Mielke
Die Bahn
michael.mielke@bahn.de

Abstract: In this paper, we address two specific areas: (a) how economic theory could provide a sound foundation to information quality (IQ) costing and (b) the extent both failure mode and effects analysis (FMEA) and service quality (SQ) literature could provide an empirical framework for estimating IQ cost curves. A case study of how risk tends to impact problem ticket requests (PTRs) and system change requests (SCRs) costing is show-cased. The empirical results show that the potential risks cannot be overlooked in so far as costing software maintenance projects.

Key Words: Economic theory, Failure Mode & Effects Analysis, Service Quality, Data Quality

INTRODUCTION

Organizations cannot improve their ability to capture and harness the power of information unless they are able to determine the value of information quality (IQ). In fact, a 100 percent customer satisfaction is a service quality (SQ) goal that can be achieved if organizations pursue managing the 5-20 percent defective data they tend to maintain at any given time [1]. Researchers have since developed clearer definitions of IQ such as accuracy, believability, relevancy, and timeliness [Ibid] – these are among the several dimensions of IQ which determine whether the quality of information meets or exceeds the requirements needed to solve business problems. Decision makers have to come to terms with the true cost of IQ as their decisions on IT investments and expectations of market performance truly depend on it.

BACKGROUND, RATIONALE, AND PURPOSE

While IQ is a vital aspect of the SQ literature, it is often treated by some organizations as a product attribute and not as a product. This in viewpoint could likely have a different ramification to the core IT business processes of a company [2]. That is, treating IQ as a product more than likely entails the involvement of specific processes such as the churning, marketing, and distribution of "IQ widgets". In addition, it remains to be seen how IQ's nominal costs tend to vary from expected opportunity costs, which are the highest benefits forgone (e.g., cost as function of risk). In other words, it is one thing to obtain software costs estimates to fix a system functionality issue. It is another to quantify cost based on risk, e.g., by determining the cost of activities through valuing the inspection of products, excessive engineering changes, the cost of doing rework, and estimating repairs of substandard equipment [3].

Another concern about IQ is that while some researchers (e.g., [4], [5], [6], [7], and [8]) speculate on the potential shape of IQ cost curves, they often lack the theoretical rigidity (e.g. economic theory) need to model IQ cost schedules.

> This approach is similar to the manner economists trace the production of widgets from the behavior of individual agents - regardless of whether they are households, firms or other organizations - are assumed to behave rationally, e.g., purposefully, and that their behavior can be described as if they maximized a specific objective function (e.g., utility). Economists tend to apply the principle of rational, optimizing behavior to areas where researchers formerly assumed that behavior is habitual and often downright irrational.

Other empirically related questions have also remained unanswered as well. For instance, does it matter to the organization if the IQ cost curve is not linear in the logs (logarithm)? How will the shape of the IQ cost curve have any bearing on public policy {e.g., optimal IQ tax scenarios on information as the growing debate on e-commerce continues ([9] and [20])}? Thus, this paper also shows that the technique called failure mode and effects analysis (FMEA) could be utilized not only to address IQ problems as a subset of service quality failures (SQFs), but also to serve as a potential tool for building IQ cost curves. Typically, SQF actions require administrative actions that include documenting and classifying complaints, creating internal complaint forms, accessing complaints made to front-line employees, and categorizing customers who complain [10]. FMEA addresses those actions by incorporating notions of risks in the costing analysis. This paper therefore addresses how: (a) IQ related costing can be better quantified and (b) to empirically estimate the shape of IQ cost curves.

Literature Review

The quality cost literature is primarily based on the Total Quality Management (TQM) movement. For instance, one aspect in the literature proposes the taxonomy of data quality costs which could be used as a foundation for data quality cost benefit considerations. Specifically, [7] query that “there is no validated economic theory of data quality costs that could be used as a basis for data quality cost analysis”, while some authors ([11] and [12]) utilize the cost-benefit analysis (CBA) literature to explain a part of that gap. In [11], the research direction takes the path of exploring the demand for information integrity, which is assumed to be driven by bounded rationality (BR) and how the agency theory concept can be interpreted using a decision making model. On the other hand, [12] develop equations used in calculating the value of information and of improvement in the value of information integrity using CBA. What is absent in these research efforts is how to evaluate DQ based on the requirement to treat it as a product [13].

Treating IQ as a Product

In [13], the authors define the creation of IQ from a manufacturing concept. Accordingly, the following can be observed about IQ:

“(a) Output: is akin to manufacturing a physical product such as raw materials, storage, assembly, processing, inspection, rework, and packaging (formatting). Common IQ output items -- management reports, invoices, dashboard indicators, etc. are assembled in a production line. The operations needed to churn components and /or processes of an IQ may be outsourced to organizations that use a different set of computing resources.

(b) Taxonomy: IQ products can be grouped based on similar characteristics and common data inputs permitting this group to be managed as a whole -- multiple Information Products (IP) may share a subset of processes and data inputs, and may be created using a single “production line” with minor variations that distinguish each IQ product;

(c) Business Process Improvement Methods: TQM methods at source, which are successfully applied in manufacturing, can be adapted for IQ. By systematically using the manufacturing stages and evaluating data quality at each stage, the implications of poor-quality data can be evaluated. Some implications for IQ include “impacts of delays in one or more manufacturing stages, trace a quality-problem in an IQ to the manufacturing stage(s) that may have caused it, and predict the IP(s) impacted by quality issues identified at some manufacturing step(s).”

Economics of Information Quality

In this paper, typical IQ problems are portrayed as in the economics literature with the use of production possibilities frontier as provided by [14] and [15]. The three questions concerning the likely production, distribution, and consumption of IQ are: how, what, and for whom? The assumption of handling IQ as if it were traded and valued in the market is hereby imposed. This crucial assumption allows the IQ researcher to comment on the treatment of IQ as a product to allow for the endogenous changes of market forces and the pricing system to affect the supply IQ curve.

The Pricing System

As in [22], organizations can be thought of as "small IQ factories" which produces IQ products such as IQ goods using time and input of ordinary market goods, "semi-manufactures", which the organization "purchases" on the "market". In this type of analysis, prices of basic goods have two components. The first is comprised of the direct costs of purchasing intermediate goods on the market. The second is the time expenditure for production and consumption of the good in question for a specific good, this time expenditure is equivalent to wages multiplied by the time spent per unit of the good produced in the organization. This implies that an increase in the wage of one IQ laborer gives rise not only to changed incentives for work on the market, but also to a shift from more to less time-intensive product on and consumption of goods produced by the organization, i.e., basic goods.

The pricing system coordinates markets so that rational agents, consumers and producers of IQ, pursuing their own self interest and end up creating a coordinated and smoothly operating economy. As IQ becomes more abundant (scarce), their prices tend to increase (decrease) resulting to consumers pursuing their own self interest, which have an incentive to start economizing (splurging) on their use of the scarce resource, and substituting other products that are now relatively cheaper (exorbitant). Simultaneously, producers pursuing their own self interest have an incentive to begin producing more (less) of the more precious IQ commodity because profits generally increase (decrease) with prices. These simultaneous actions by consumers and producers of IQ are the appropriate responses when an IQ product begins to become more precious (less preferred). Thus, such a pricing mechanism is what coordinates consumers' and producers' rational maximizing behavior.

How: Non-market valuation

IQ is usually attributed as a key feature of information products, and not as marketed products. However, the real price of IQ can be estimated either in the form of willingness to pay for a price change or attribute change. Figures 1 and 2 depict how the demand curve is derived through equivalent variation and through willingness to pay (WTP). The mathematical derivation for these charts is specified in [16]. In WTP models, estimation methods such as contingent valuation method (CVM) are drafted based on individuals’ hypothetical behavior on markets that are set up for IQ in a survey setting. CVM is a widely used method in the field of externalities (incidental outcomes of legitimate economic activities which are not usually traded in the market place) such as providing estimates of elasticities of demand and willingness to pay WTP for environmental services. If the decision maker's dilemma stems from how to treat the nominal value (actual cost or benefit) of a public investment – e.g., it may not reflect the real investment costs or benefits to society, shadow pricing comes into play as well.

> The *Economist* definition of shadow pricing is that which pertains to "an opportunity cost of an activity which can be calculated by capturing all the variables involved in a decision and not merely those for which market prices exist" [18]. In [21], the authors warn that there is a misconception in the public that the Internet is free – in fact, research shows that "while the marginal cost for Internet traffic may approach zero due to statistical sharing, other costs, such as congestion costs, may be significant". Other researchers [21] find that in network optimization problems, "classes of algorithms can be developed and interpreted in terms of either congestion indication feedback signals or explicit rates based on shadow prices". As economics theory suggests, IQ decisions on what to produce, what price to charge, and how much to produce will hinge on how much people are willing to buy and sell IQ at different prices.

Figure 1. Compensating and Equivalent Variations

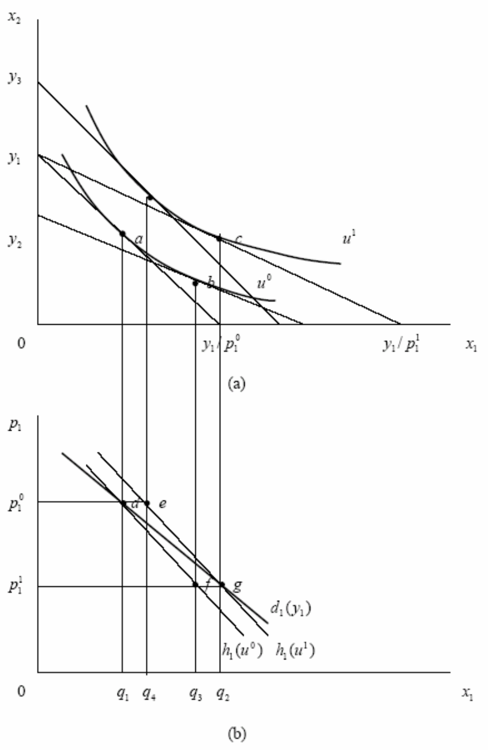
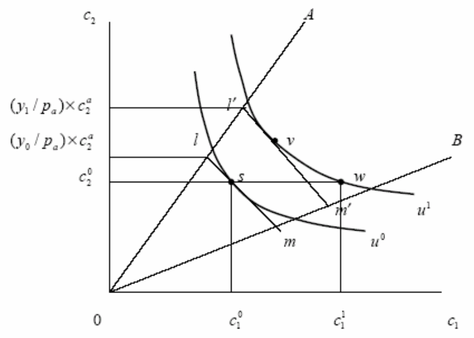


Figure 2. Willingness-to-Pay Model for an Attribute Change



What: Demand for IQ

Consumers of IQ need to act on their purchasing decisions, which are based on the price of the product P, price of substitutes P_S, price of complements P_C, future expected prices P^e, tastes T, income Y, and population N, which are expressed in the function: $Q^D = f(P, P_S, P_C, P^e, T, Y, N)$. The mathematical derivation of Q^D is not provided in this paper, but is referenced in [15]. Holding all other variables constant with exception to price (*ceteris paribus*), the demand function becomes: $Q^D = f(P)$. Figure 3 portrays the hypothetical graph of demand expressed as the inverse relationship of price, $P = f^{-1}(Q^D) = 10 - Q^D$.

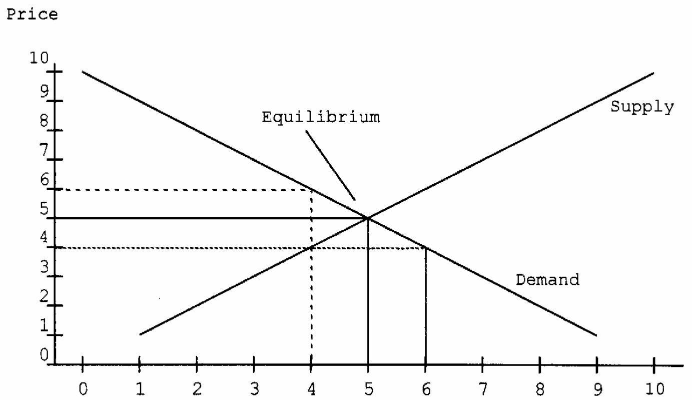


Figure 3: Hypothetical Demand and Supply for Data Quality

For Whom: Supply of IQ

Producers of IQ need to act on purchasing decisions based on price of the product P, technology T, price of inputs P_i, price of substitutes P_S (in production), price of complements P_C (in production), future expected prices P^e, number of firms F, goals of the firm G, which is expressed in the function: $Q^S = g(P, T, P_i, P_S, P_C, P^e, F, G)$. The mathematical derivation of Q^S is not provided in this paper, but is cited in [15]. Salient point: Supply curves stem from a flow concept, which is the amount of IQ supplied per unit of time, a schedule; *ceteris paribus*, generally $Q = f(P)$. Thus, this relationship will be positive: more IQ quantity will be supplied at higher prices than at lower prices.

Equilibrium

The equilibrium price and quantity where S = D is the point where there are no forces that will change the dynamics. The equilibrium is stable as these forces will automatically slip back any displacement of S and D back to a new equilibrium position.

To summarize, the economics of information quality hinges upon the treatment of IQ as a product. Once that initial premise has been assumed, the analyst is able to take the mathematical derivation of supply and demand curves for IQ and speculate on the dynamics of those curves as they are impacted by the decision maker's policies, market conditions, or public policies.

METHODS

As identified by [10], companies that address SQ need to "generate additional information on service quality, disseminate it to those responsible for implementing improvements, and identify those process improvements that will have the greatest impact on profitability." These SQ efforts are also directly related with FMEA, a systems engineering process that identifies the likelihood of errors in business processes, and allows for process improvement over time ([19] and [20]). The common denominator between the SQ and FMEA literature are complimentary processes that lead to the identification of the components of the total cost of IQ (TCIQ). As adopted from [7] and [8], the total cost components of IQ can be divided into three areas: Prevention + Appraisal + Failure Costs = TCIQ.

> While the detection of data defects is possible in spite of corrections, efforts to reconcile of data and continuing measurements to prevent errors from happing again are needed -- these efforts are called prevention/repair/detection costs. To prevent poor quality, coding errors, design errors, mistakes in the user manuals, as well as badly documented or flawed complex codes must be corrected through expenditures in the programming, design, and marketing departments of, say, a software company. For example by applying business rules (data constraints) some, but not all data input errors can be automatically detected and corrected.
>
> Appraisal costs are costs of activities designed to find defects in the system such as code inspections and through various types of testing. Design reviews are part prevention and part appraisal in that looking for errors in the proposed design itself during the review can be accomplished during the appraisal -- looking for ways to strengthen the design implies prevention as well.
>
> Failure costs are costs that result from poor quality such as the cost of fixing bugs and the cost of dealing with customer complaints. The two types of failure costs stem from those that arise before and after the product is delivered to the customer. These costs stem from both internal and external failure costs. Internal costs are borne by groups outside of the product development phase -- the costs of the wasted time, the missed milestones, and the overtime to get back onto schedule are all internal failure costs. External Failure Costs are costs that arise after the product has been delivered to the customer, such as customer service costs, or the cost of patching a released software product and distributing the patch.

During ICIQ 2004, it was proposed that the costs of problem ticket requests (PTRs) and system change requests (SCRs) are likely are likely to be biased ([19] and [20]) if risk, in the form of FMEA is not utilized in the costing analysis. PTRs and SCRs are being utilized to capture data quality problems during systems integration and IT production. There are three possible ways to estimate bias: (a) by taking the product of nominal costs with the probability of fault and the probability that it escapes detection; (b) by simply taking the product between nominal costs with the weighted risk priority number (RPN); and (c) through statistical analysis of omitted variables. Suppose that p_f and p_d are independent vectors of probability underlying failure and non-detection, respectively, and that the probability that the user receives the IQ problem or defect is equal to the product of both risk variables, $p_f * p_{d.}$ Thus, if n items are produced periodically, then the expected cost EC of the PTR and SCR is equal to $C_n * p_f * p_d$., where C_n is the original nominal cost (ONC) multiplied by the frequency of occurrence or the total number work-arounds in a given period. The difference between EC and the original nominal costs will determine how far the real costs are divergent from nominal costs. The third way to estimate bias is to apply Ordinary Least Squares (OLS) on ONC and by using RPN and DURATION as the key regressors. If a key regressor's coefficient is insignificant (significant), then, it likely does not (does) contribute to the explained variations of the model. Thus, omitting best, linear, and unbiased predictors of the model such RPN or DURATION would likely bias the results.

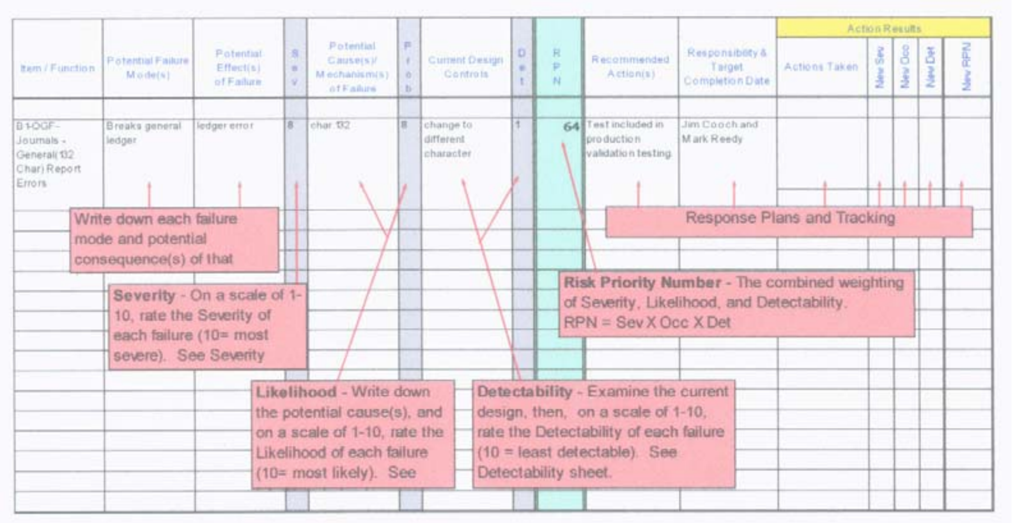
The Data

A panel data of 3,938 samples were obtained for ONC, RPN, and DURATION. They were collected from an Oracle local development Department of Defense database of PTRs and SCRs for the period July 2003 through January 2005 (Figure 4 illustrates sample data). ONC (which are near estimates of costs and not actual costs were utilized to preserve procurement integrity) increases as a function of vendor level of effort or VLOE. The VLOE rate used in this paper is $1200 per day; however, this rate can be altered during sensitivity analysis. RPN is a vector of scores, which is usually derived from multiplying the probabilities of fault and non-detection with the system FMEA severity rating (Figure 5). These RPN scores stem from the weekly decisions of Government process leads who conduct relative pair-wise comparisons and rankings of SCRs and PTRs. Each RPN category's minimum is 10, while its maximum is 10^3 or 1000. DURATION is simply the time in number of days it takes for the vendor to accomplish the PTR / SCR.

Oracle Local Development Work in Progress (04/12/04)

	TAR#	Description	Duration	Start	Finish	Need By	Phase Ne	Priority	Owner	Status	Develope
1	DESC1983	Returns Without Credit	1 day?	03/18/04	03/18/04	04/01/04	PC&S-B E	1000	Davis, Pat	BUS	Earl
2	DESC2516/F1	SIP/Price Series Code Upload	1 day	01/23/04	01/23/04	04/01/04	PC&S-B E	1000	Thompson	DEV	Earl
3	DESC2914	EDI Rollup/Breakout Process Change for Tax Rate	1 day?	03/30/04	03/30/04	04/30/04	PC&S-B E	1000	Coffel, Bet	LOG	Julius
4	DESC2785	Modification to the Contract Activity Report	5 days	02/09/04	02/13/04	02/27/04	Bulk Prod	600	Comar, Bil	LOG	Julius
5	DESC2844	Modification to the Contract Activity Report	3 days	02/04/04	02/06/04	03/01/04	Bulk Prod	600	Comar, Bil	LOG	Julius
6	DESC2907	Add Clin to Contract Activity Disbursement Report	1 day?	03/29/04	03/29/04	04/30/04	Production	600	Coffel, Bet	LOG	Julius
7	DESC2911	Timeouts and Return Values for Custom Programs	1 day?	03/30/04	03/30/04	04/20/04	Production	600	Smith, Ma	LOG	Earl
8	DESC2651	Lowest Projected Inventory Includes Estimates	10 days	01/19/04	01/30/04	09/30/03	Bulk Prod	470	Barnett, Cl	LOG	Dave/Tracy
9	DESC1868	OED to FES interface validation Super User Book Inventory V	10 days	01/08/04	01/21/04	NA	Production	400	Cerda, Joh	LOG	Wally
10	DESC2852	Retiring of scripts	1 day	02/09/04	02/09/04	NA		400	Smith, Ma	LOG	Julius
11	DESC2728	OED Tanker Lift Schedule reports adding Lift Area	4 days	02/04/04	02/09/04	01/07/04	Bulk Prod	256	Barnett, Cl	LOG	Dave
12	DESC2722	BargeNet Report adding Sort parameter	2 days	02/02/04	02/03/04	01/07/04	Bulk Prod	255	Barnett, Cl	LOG	Dave
13	DESC2851	Issues Consumption Chart using Loc of Mvt Id vs Load L	2 days	02/09/04	02/10/04	02/09/04	Bulk Prod	235	Barnett, Cl	LOG	Tracy
14	DESC2397	Weighted Average Price (waiting on requirements)	0 days	01/22/04	01/22/04	03/31/04	Bulk Prod	230	Todd, Bart	BUS	Tracy
15	DESC2883	Transaction by Mvt-Disc Report	1 day?	03/10/04	03/10/04	03/10/04		230	Brooks, Al	LOG	Tracy
16	DESC2802	Need to Re-establish the load process from DAAS to the DOD	10 days	03/22/04	04/02/04	NA		225	Tolbert, Ri	LOG	Earl
17	DESC2841	Update DAAS Master File for Sales	5 days	02/04/04	02/10/04	05/15/04		225	Brooks, Al	LOG	Earl
18	DESC2908	Improve Inv Compare to Guides Upload	1 day?	03/29/04	03/29/04	05/03/04		220	Barnett, Cl	LOG	Dave
19	DESC2681	Daily Inventory and Movements Improvements for book inv, rec	5 days	02/18/04	02/24/04	01/29/04	Bulk Prod	160	Barnett, Cl	LOG	Dave
20	DESC2301	Forecasting Sales for the Projected Inventory Report	1 day?	02/10/04	02/10/04	12/05/03	Bulk Prod	155	Barnett, Cl	LOG	Dave
21	DESC2687	Distribution System Projected Inventory Sheet Type	15 days	02/25/04	03/16/04	01/07/04	Bulk Prod	150	Barnett, Cl	LOG	Dave
22	DESC2694	OED Interface to run Overdue Tanker Moves Report	1 day?	03/17/04	03/17/04	02/02/04		150	Brooks, Al	LOG	Dave
23	DESC2739	New S/H version of the Unmatched Stock Transfers Report	3 days	01/30/04	02/03/04	02/02/04		150	Brooks, Al	LOG	Tracy
24	DESC2740	New S/H version of the Unmatched Purchases Report	3 days	02/04/04	02/06/04	02/02/04		150	Brooks, Al	LOG	Tracy
25	DESC2468	Redwood Report Repository Process Improvments	10 days	02/16/04	02/27/04	09/30/03		140	Weber, Jo	LOG	Julius
26	DESC2753	Multiple Combined Liability Report	2 days	02/10/04	02/11/04	03/15/04		140	Brooks, Al	DEV	Tracy
27	DESC2834	Report Gating History Report	3 days	01/29/04	02/02/04	01/30/04	Production	140	Brooks, Al	LOG	Tracy
28	DESC2821	Modification to Remittance Address batch job	3 days	02/04/04	02/06/04	03/15/04	Production	138	Comar, Bil	LOG	Earl
29	DESC2830	Quarterly Sales Quarter Definitions and Cross tab re	3 days	01/22/04	01/26/04	03/15/04		137	Barnett, Cl	LOG	Tracy
30	DESC2689	Issues Consumption Chart add Region and Distn Sys	2 days	02/12/04	02/13/04	02/11/04	Bulk Prod	135	Barnett, Cl	LOG	Tracy
31	DESC2896	Modification to Fuel Receipts Summary Detail tab	1 day?	03/22/04	03/22/04	NA		134	Barnett, Cl	LOG	Tracy

Figure 1: From gap analysis, rank failure mode by risk priority number score



											Action Results				
Item / Function	Potential Failure Mode(s)	Potential Effect(s) of Failure	Sev	Potential Cause(s)/ Mechanism(s) of Failure	Prob	Current Design Controls	Det	RPN	Recommended Action(s)	Responsibility & Target Completion Date	Actions Taken	New Sev	New Occ	New Det	New RPN
B1-OGF-Journals - General(132 Char) Report Errors	Breaks general ledger	ledger error	8	char 132	8	change to different character	1	64	Test included in production validation testing.	Jim Cooch and Mark Reedy					

Write down each failure mode and potential consequence(s) of that

Response Plans and Tracking

Severity - On a scale of 1-10, rate the Severity of each failure (10= most severe). See Severity

Risk Priority Number - The combined weighting of Severity, Likelihood, and Detectability.
RPN = Sev X Occ X Det

Likelihood - Write down the potential cause(s), and on a scale of 1-10, rate the Likelihood of each failure (10= most likely). See

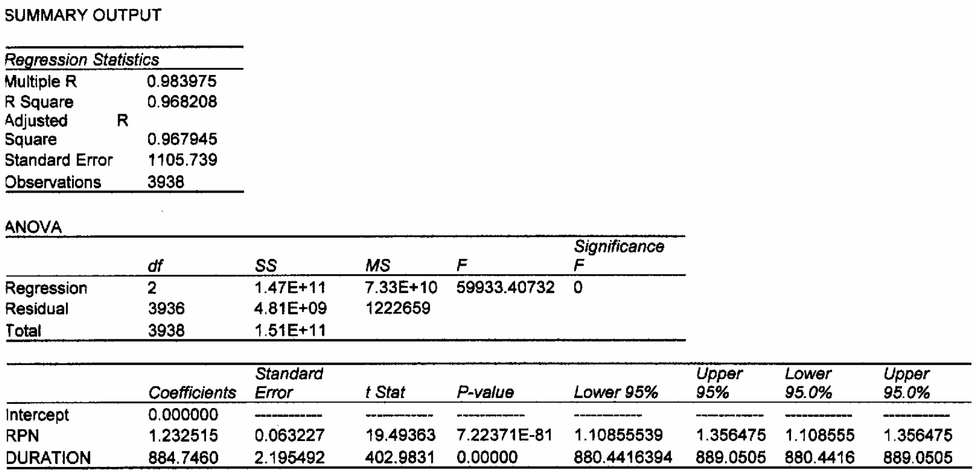
Detectability - Examine the current design, then, on a scale of 1-10, rate the Detectability of each failure (10 = least detectable). See Detectability sheet.

Figure 2: FMEA Tableau Procedures

RESULTS

Applying OLS on ONC, by way of the regressors RPN and DURATION provides the estimated statistical model $\text{ONC}_{\text{HAT}} = f\{\text{RPN}_{\text{HAT}}, \text{DURATION}_{\text{HAT}}, \text{and } \hat{E}\}$, where ONC_{HAT}, RPN_{HAT}, and $\text{DURATION}_{\text{HAT}}$ are statistically derived coefficients from regressing PTR/SCR cost on risk priority number and the completion time in days to fix the problem or include a missing IT functionality. $\hat{E}$ represents the residual or the stochastic error term of the equation, which is the portion of the model is that is unexplained by the regressors.

Figure 6 show that both RPN and DURATION likely explains 96 percent of the variation in ONC in the sample. The coefficients show that a one percent increase in RPN and DURATION will increase ONC by $1.24 and $884.75, respectively. Both the positive signs and individual coefficients are significant at the 95 percent level. Interestingly, if the impact of RPN on ONC is not taken into account as in the case of an omitted variable, it is estimated that ONC could be biased by about 5 percent of the time, or about plus or minus $953,047, from 2003 through the first quarter of 2005.



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.983975
R Square	0.968208
Adjusted R Square	0.967945
Standard Error	1105.739
Observations	3938

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1.47E+11	7.33E+10	59933.40732	0
Residual	3936	4.81E+09	1222659		
Total	3938	1.51E+11			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.000000	------------	------------	------------	------------	------------	------------	------------
RPN	1.232515	0.063227	19.49363	7.22371E-81	1.10855539	1.356475	1.108555	1.356475
DURATION	884.7460	2.195492	402.9831	0.00000	880.4416394	889.0505	880.4416	889.0505

Figure 3: Results from Regressing PTR/SCR Original Nominal Cost on Risk Priority Number and Duration

DISCUSSION

First, the model results primarily show that while about 5 percent of the variations in ONC are explained by RPN, its significance and relevance cannot be ignored. Thus, these results stress the important role of maintaining a strong applications support group (ASG), a body that would press on collecting statistics on the cost of rework such as software internal scrap data, rework, warranty costs, other cost-of-poor-quality factors, proxies for data quality failure work-arounds, and other systems aspects such as incidents of downtime. Ultimately, estimates of cost per fault EC can be multiplied with both the number of items and the probabilities of fault and no detection to see whether they truly create an expected cost bias in the same magnitude as stated above.

Second, TCIQ is defined as a function of the variables prevention, appraisal, and failure costs. Interestingly, these categories of cost closely mesh with those of the categories of risk that are itemized in FMEA. From Figure 5, FMEA is attributed to: (a) potential failure modes and potential effects of failure are captured under severity; (b) potential appraisal and causes/mechanisms of failure are under the probability of failure; and (c) the manner current design and prevention controls are captured under the probability of non-detection. It is therefore proposed that the model $\text{ONC}_{\text{HAT}} = \text{RPN}_{\text{HAT}}*\text{RPN} + \text{DURATION}_{\text{HAT}}*\text{DURATION} + \hat{E}$ likely approximates the theoretical cost profile stated in [7] and [8] (also see DQ cost taxonomy in Figure 7). In other words, it is suggested that ONC, which is significantly correlated with VLOE, and is likely broken down by prevention, appraisal, and failure costs.

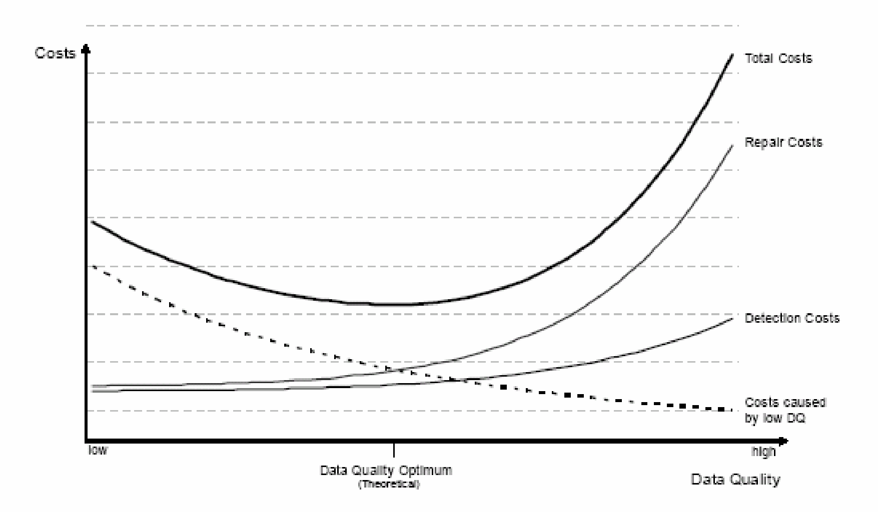


Figure 4: Theoretical Supply Curve for DQ (Eppler and Helfert, 2004)

A graphic representation (scatter plot of the log of actual versus predicted costs) of the model is presented in Figure 8. The data was obtained by taking the logarithmic transformation of both sides of the model before applying OLS. The results of this regression is compatible with the economic theory pertaining to supply curves in so far as the slope (elasticity of the IQ supply cost curve with respect to the variables) is concerned as it is mathematically linear in the logs: $\text{LOG}\{\text{ONC}\}_{\text{HAT}} = \text{ß} + \text{ß}_1*\text{LOG}\{\text{RPN}\}_{\text{HAT}} + \text{ß}_2*\text{LOG}\{\text{DURATION}\}_{\text{HAT}} + \hat{E}$. This version of the model (Figure 9) shows that a 1% change in RPN will likely increase ONC by 0.02 percent, while a 1% change in DURATION will increase it by about 1 percent. In addition, a 1% change in the intercept ß will increase ONC by 3.1 percent. All coefficient estimates are significant at the 95% confidence level.

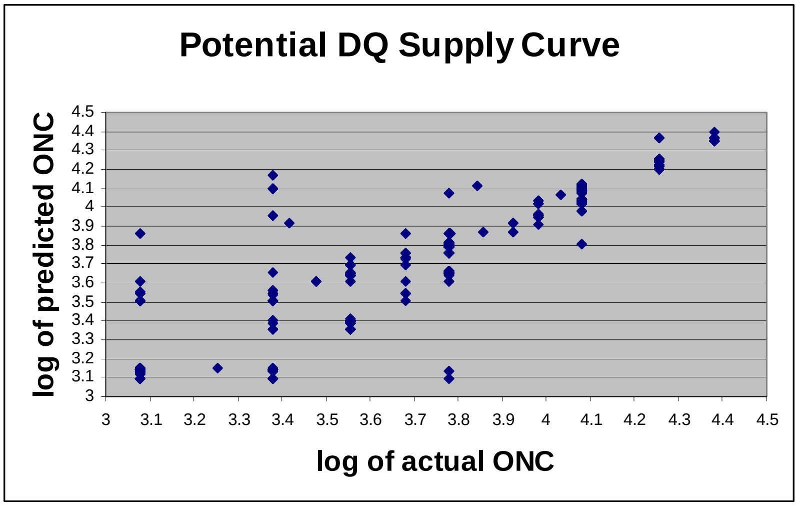
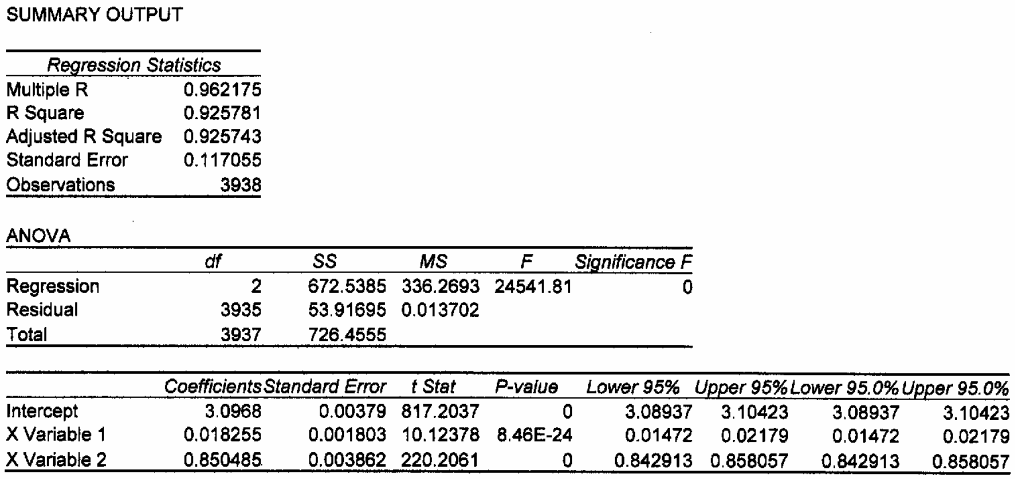


Figure 5: Scatter plot of the log of actual nominal cost vs. predicted values of ONC



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.962175
R Square	0.925781
Adjusted R Square	0.925743
Standard Error	0.117055
Observations	3938

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	672.5385	336.2693	24541.81	0
Residual	3935	53.91695	0.013702		
Total	3937	726.4555			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.0968	0.00379	817.2037	0	3.08937	3.10423	3.08937	3.10423
X Variable 1	0.018255	0.001803	10.12378	8.46E-24	0.01472	0.02179	0.01472	0.02179
X Variable 2	0.850485	0.003862	220.2061	0	0.842913	0.858057	0.842913	0.858057

Figure 6: Results from applying Ordinary Least Squares on the logs of DEP and INDEP variables

LIMITATIONS

The paper has a few limitations that can be addressed in future work. First, it utilizes economic theory to pursue the rationale that supply schedules can estimated by way of market and non-market valuations. The problem with market valuation is that it assumes that DQ is traded in the market and that there exists some sort of market clearing price for DQ. Clearly, this is assumption is far from reality, although this pushes the notion that if DQ is to be treated as a product, it must be rationalized as a theoretical widget. On the other hand, the idea of estimating supply curves based on non-market valuation of DQ may seem feasible, yet this field of research is likely to be at its infancy stage as willingness-to-pay economic surveys for DQ still remains to be seen. The second limitation of the paper is the stylized empirical model, which is derived from two assumptions. First, it proposes that the DQ supply curve itself is the summation of Prevention, Appraisal, and Failure Costs as validated by [5], [6], [7], and [8]. Then, the paper proposes that FMEA methods (severity, likelihood of failure, detectability, and risk priority) likely approximate or lead to risk-based costs valuation. More research, explanation, or examples could be utilized to make this connection. Perhaps as addressed by [8], the impact of introducing prevention measures on detection and repair costs might be the real issue. Thirdly, the empirically estimated model $\text{ONC}_{\text{HAT}} = f$ { ß, ß$_1$*LOG{RPN}$_{\text{HAT}}$, ß$_2$*LOG{DURATION}$_{\text{HAT}}$, and $\hat{E}$ } is so simple that: (a) the results could mask any other underlying drivers of costs. In other words, structural dummy variables could be introduced into the panel data as additional regressors to see if other extraneous drivers could be singled out such as the imposition of a regulatory mandates (Clinger-Cohen Act-related IT actions, “Blue Book Certification” or Federal Financial Management Improvement Act-related testing, etc.) and (b) the financial impact ($953,047) of omitting the risk variable RPN needs to be treated notionally – for the sake of procurement integrity, near estimates of costs were utilized.

CONCLUSION

This paper addresses two specific areas: (a) how economic theory is employed by the decision maker to make decisions on IT investments based on the new found interpretations of IQ and (b) the extent FMEA and the service quality literature address the foundation for IQ cost curves. This paper finds that the stylized model shows that risk cannot be overlooked in so far as accounting for the financial bias in PTR / SCR costing. It also finds that IQ cost curves are linear in the logs, thus enabling the decision maker to look at optimal levels of IQ quantities and respective pricing. Finally, if further research proves the robustness of empirically derived IQ cost curves, this will likely have implications for both private and public policy decision makers.

REFERENCES

[1] Smith, Nancy Duvergne. “Is Corporate IQ (Information Quality) High Enough? The Perils of Dirty Data Inhibit Strategic Decisions: Interview with Richard Wang, Director of MIT Information Quality (IQ)”. *Impact*, winter 2003: Faces of Research. http://web.mit.edu/ctpid/www/i11/impact11_prof.html

[2] Richard Y. Wang, Yang W. Lee, Leo L. Pipino and Diane M. Strong. "Manage Your Information as a Product". *Sloan Management Review*, summer 1998, Vol. 39, No. 4, pp. 95–105.

[3] Lawrence P. Carr. "Applying Cost of Quality to a Service Business". *MIT Sloan Management Review*, summer1992, Vol. 33, No. 4, pp. 72–77)

[4] Juran, J. M., & Gryna, F. M. (Eds.) (1988). *The Quality Control Handbook* (4th ed.). New York: McGraw-Hill.

[5] Feigenbaum, A.V. (1991, 3rd Ed. Revised), *Total Quality Control*, McGraw-Hill, Chapter 7.

[6] Feigenbaum, A.V. Total Quality Control, *Harvard Business Review*, 1956, 34(6), pp. 96-101

[7] Kaner, Cem. “Quality Cost Analysis: Benefits and Risks”. http://www.kaner.com/qualcost.htm

[8] Eppler, Martin J. and Markus Helfert. “A Framework For The Classification Of Data Quality Costs And An Analysis Of Their Progression”. The 9th International Conference on Information Quality, MIT, Cambridge, Massachusetts, 2004. 311-325.

[9] Weidenbaum, Murray. "The Fundamental Internet Tax Debate". *The Washington Quarterly*. Vol. 24, Issue 1 - Winter 2001, pp. 41 – 52 .

[10] Tax, Stephen S. and Stephen W. Brown. "Recovering and Learning from Service Failure". 1998. *Sloan Management Review*, fall 1998, Vol. 40, No. 1, pp. 75–88.

[11] Tallberg A., *An Economic Framework for Information Integrity*. Forskningsrapporter Research Reports, Swedish School of Economics and Business Administration, Finland, 1999.

[12] Mandke, Vijay and Nayar, Madhavan, “Modeling Information Flow for Integrity Analysis”. The 3rd International Conference on Information Quality, MIT, Cambridge, Massachusetts, 1999.

[13] Shankaranarayan. G., Ziad, Mostapha, and Richard Y. Wang. “Managing Data Quality in Dynamic Decision Environments: An Information Product Approach”. *Data Management*, Feb 2003.

[14] Gates, William. Various notes in economics. http://web.nps.navy.mil/~facvita/vita_gates.pdf

[15] Varian, Hall. *Intermediate Microeconomics: A Modern Approach*. Michigan: W.W. Norton & Company. 1987.

[16] Kyung Hee Lee. “Willingness to Pay for Information: An Analyst's Guide”. Department of Economics, University of Georgia. 1993

[17] Lee W. McKnight and Joseph P. Bailey. An Introduction to Internet Economics: Presented at MIT Workshop on Internet Economics March 1995. *The Journal of Electronic Publishing*. May, 1996 Volume 2, Issue 1, ISSN 1080-2711.

[18] The Economist. http://economist.com/research/Economics/alphabetic.cfm?LETTER=S

[19] Raneses, Anton. "Beyond Business Process Reengineering: Data Quality Engineering". 9th International Conference on Information Quality. The 9th International Conference on Information Quality, MIT, Cambridge, Massachusetts, 2004.

[20] Carbone, Thomas and Donald D. Tippet. “Project Risk Management Using the Project Risk Failure Mode and Effects Analysis”. *Engineering Management Journal*, volume 16, no. 4, December 2004.

[21] FP Kelly, AK Maulloo and DKH Tan. “Rate control for communication networks: shadow prices, proportional fairness and stability.” *Journal of the Operational Research Society*, 49 (1998), 237-252.

[22] Royal Swedish Academy of Sciences. http://nobelprize.org/economics/laureates/1992/press.html