Balanced Scorecards to Measure Enterprise Risk Performance

10

Balanced scorecards are one of a number of quantitative tools available to support risk planning.[1] Olhager and Wikner[2] reviewed a number of production planning and control tools, where scorecards are deemed as the most successful approach in production planning and control performance measurement. Various forms of scorecards, e.g., company-configured scorecards and/or strategic scorecards, have been suggested to build into the business decision support system or expert system in order to monitor the performance of the enterprise in the strategic decision analysis.[3] This chapter demonstrates the value of balanced scorecards with a case from a bank operation.

While risk needs to be managed, taking risks is fundamental to doing business. Profit by necessity requires accepting some risk.[4] ERM provides tools to rationally manage these risks. Scorecards have been successfully associated with risk management at Mobil, Chrysler, the U.S. Army, and numerous other organizations.[5] It also has been applied to the financial analysis of banks.[6]

Enterprise risk management (ERM) provides the methods and processes used by business institutions to manage all risks and seize opportunities to achieve their objectives. ERM began with a focus on financial risk, but has expended its focus to accounting as well as all aspects of organizational operations in the past decade. Enterprise risk can include a variety of factors with potential impact on an organizations activities, processes, and resources. External factors can result from economic change, financial market developments, and dangers arising in political, legal, technological, and demographic environments. Most of these are beyond the control of a given organization, although organizations can prepare and protect themselves in time-honored ways. Internal risks include human error, fraud, systems failure, disrupted production, and other risks. Often systems are assumed to be in place to detect and control risk, but inaccurate numbers are generated for various reasons.[7]

ERM brings a systemic approach to risk management. This systemic approach provides more systematic and complete coverage of risks (far beyond financial risk, for instance). ERM provides a framework to define risk responsibilities, and a need to monitor and measure these risks. That's where balanced scorecards provide a natural fit—measurement of risks that are key to the organization.

D.L. Olson, D.D. Wu, *Enterprise Risk Management Models*, Springer Texts in Business and Economics, DOI 10.1007/978-3-662-53785-5_10

ERM and Balanced Scorecards

Beasley et al.[8] argued that balanced scorecards broaden the perspective of enterprise risk management. While many firms focus on Sarbanes-Oxley compliance, there is a need to consider strategic, market, and reputation risks as well. Balanced scorecards explicitly link risk management to strategic performance. To demonstrate this, Beasley et al. provided an example balanced scorecard for supply chain management, outlined in Table 10.1.

Other examples of balanced scorecard use have been presented as well, as tools providing measurement on a broader, strategic perspective. For instance, balanced scorecards have been applied to internal auditing in accounting[9] and to mental health governance.[10] Janssen et al.[11]applied a system dynamics model to the marketing of natural gas vehicles, considering the perspective of sixteen stakeholders ranging across automobile manufacturers and customers to the natural gas industry and government. Policy options were compared, using balanced scorecards with the following strategic categories of analysis:

- Natural gas vehicle subsidies
- Fueling station subsidies
- Compressed natural gas tax reductions
- Natural gas vehicle advertising effectiveness.

Balanced scorecards provided a systematic focus on strategic issues, allowing the analysts to examine the nonlinear responses of policy options as modeled with system dynamics. Five indicators were proposed to measure progress of market penetration:

1. Ratio of natural gas vehicles per compress natural gas fueling stations
2. Type coverages (how many different natural gas vehicle types were available)
3. Natural gas vehicle investment pay-back time
4. Sales per type
5. Subsidies par automobile

Small Business Scorecard Analysis

This section discusses computational results on various scorecard performances currently being used in a large bank to evaluate loans to small businesses. This bank uses various ERM performance measures to validate a small business scorecard (SBB). Because scorecards have a tendency to deteriorate over time, it is appropriate to examine how well they are performing and to examine any possible changes in the scoring population. A number of statistics and analyses will be employed to determine if the scorecard is still effective.

Table 10.1 Supply chain management balanced scorecard

Measure	Goals	Measures
Learning & growth for employees To achieve our vision, how will we sustain our ability to change & improve?	Increase employee ownership over process	Employee survey scores
	Improve information flows across supply chain stages	Changes in information reports, frequencies across supply chain partners
	Increase employee identification of potential supply chain disruptions	Comparison of actual disruptions with reports about drivers of potential disruptions
	Risk-related goals:	
	Increase employee awareness of supply chain risks	Number of employees attending risk management training
	Increase supplier accountabilities for disruptions	Supplier contract provisions addressing risk management accountability & penalties
	Increase employee awareness of integration of supply chain and other enterprise risks	Number of departments participating in supply chain risk identification & assessment workshops
Internal business processes To satisfy our stakeholders and customers, where must we excel in our business processes?	Reduce waste generated across the supply chain	Pounds of scrap
	Shorten time from start to finish	Time from raw material purchase to product/service delivery to customer
	Achieve unit cost reductions	Unit costs per product/service delivered, % of target costs achieved
	Risk-related goals:	
	Reduce probability and impact of threats to supply chain processes	Number of employees attending risk management training
	Identify specific tolerances for key supply chain processes	Number of process variances exceeding specified acceptable risk tolerances
	Reduce number of exchanges of supply chain risks to other enterprise processes	Extent of risks realized in other functions from supply chain process risk drivers
Customer satisfaction To achieve our vision, how should we appear to our customers?	Improve product/service quality	Number of customer contact points
	Improve timeliness of product/service delivery	Time from customer order to delivery
	Improve customer perception of value	Customer scores of value
	Risk-related goals:	
	Reduce customer defections	Number of customers retained
	Monitor threats to product/service reputation	Extent of negative coverage in business press of quality

(continued)

Table 10.1 (continued)

Measure	Goals	Measures
	Increase customer feedback	Number of completed customer surveys about delivery comparisons to other providers
Financial performance To succeed financially, how should we appear to our stakeholders?	Higher profit margins	Profit margin by supply chain partner
	Improved cash flows	Net cash generated over supply chain
	Revenue growth	Increase in number of customers & sales per customer; % annual return on supply chain assets
	Risk-related goals:	
	Reduce threats from price competition	Number of customer defections due to price
	Reduce cost overruns	Surcharges paid, holding costs incurred, overtime charges applied
	Reduce costs outside the supply chain from supply chain processes	Warranty claims incurred, legal costs paid, sales returns processed

Developed from Beasley et al. (2006)

ERM Performance Measurement

Some performance measures for enterprise risk modeling are reviewed in this section. They are used to determine the relative effectiveness of the scorecards. More details are given in our work published elsewhere.[12] There are four measures reviewed: the Divergence, Kolmogorov-Smirnov (KS) Statistic, Lorenz Curve and the Population stability index. **Divergence** is calculated as the squared difference between the mean score of good and bad accounts divided by their average variance. The dispersion of the data about the means is captured by the variances in the denominator. The divergence will be lower if the variance is high. A high divergence value indicates the score is able to differentiate between good and bad accounts. Divergence is a relative measure and should be compared to other measures. The KS Statistic is the maximum difference between the cumulative percentage of goods and cumulative percentage of bads for the population rank-ordered according to its score. A high KS value shows it is very possible that good applicants can receive high scores and bad applicants receive low scores. The maximum possible K-S statistic is unity. **Lorenz Curve** is the graph that depicts the power of a model capturing bad accounts relative to the entire population. Usually, three curves are depicted: a piecewise curve representing the perfect model which captures all the bads in the lowest scores range of the model, the random line as a point of reference indicating no predictive ability, and the curve lying between these two capturing the discriminant power of the model under evaluation. **Population stability index** measures a change in score distributions by comparing the frequencies of the corresponding scorebands, i.e., it measures the difference between two populations. In practice, one can judge there is no real change between the populations if an index value is no larger than and a definite population change if index value is greater than 0.25. An index value between 0.10 and 0.25 indicates some shift.

Data

Data are collected from the bank's internal database. 'Bad' accounts are defined into two types: 'Bad 1' indicating Overlimit at month-end, and 'Bad 2' referring to those with 35 days since last deposit at month-end. All non-bad accounts will be classified as 'Good'. We split the population according to Credit Limit: one for Credit Limit less than or equal to $50,0000 and the other for Credit Limit between $50,000 and $100,000. Data are gathered from two time slots: observed time slot and validated time slot. Two sets (denoted as Set1 and Set2) are used in the validation. Observed time slots are from August 2002 to January 2003 for Set1 and from September 2001 to February 2002 for Set2 respectively. While this data is relative dated, the system demonstrated using this data is still in use, as the bank has found it stable, and they feel that there is a high cost in switching. Validated time slot are from February 2003 to June 2003 for Set1 and from March 2002 to July 2002 for Set2 respectively. All accounts are scored on the last business day of each month. All non-scored accounts will be excluded from the analyses.

Table 10.2 gives the bad rates summary by Line Size for both sets while Table 10.3 reports the score distribution for both sets, to include the Beacon score accounts. From Table 10.2, we can see that in both sets, although the number of Bad1 accounts is a bit less than that of Bad2 accounts, it is still a pretty balanced

Table 10.2 Bad loan rates by loan size

Limit	Bad loans 1 Jan. 2003 (set1)			Bad loans 2 Jan. 2003 (set1)		
	N	# of bad loans	Bad rate (%)	*N*	# of bad loans	Bad rate (%)
≤$50 M	59,332	5022	8.46	61,067	1127	1.85
$50–100 M	6777	545	8.04	7000	69	0.99
Total	66,109	5567	8.42	68,067	1196	1.76
	Bad loans 1 Feb. 2002 (set2)			Bad loans 2 Feb. 2002 (set2)		
	N	# of bad loans	Bad rate (%)	*N*	# of bad loans	Bad rate (%)
≤$50 M	61,183	5790	9.46	63,981	1791	2.80
$50–$ 100 M	6915	637	9.21	7210	88	1.22
Total	68,098	6427	9.44	71,191	1879	2.64

Note: Bad 1: Overlimit; Bad 2: 35+ days since last deposit and overlimit

Table 10.3 Score statistical summary

Score band	Bad loans 1 Jan. 2003 (set1)			Bad loans 2 Jan. 2003 (set1)		
	N	Bad	Bad rate (%)	*N*	Bad	Bad rate (%)
0	1210	125	10.33	1263	27	2.14
1–500	152	58	38.16	197	27	13.70
501–550	418	117	27.99	508	49	9.65
551–600	1438	350	24.34	1593	109	6.84
601–650	4514	858	19.01	4841	194	4.01
651–700	11,080	1494	13.48	11,599	321	2.77
701–750	18,328	1540	8.40	18,799	312	1.66
751–800	21,083	888	4.20	21,356	149	0.70
≥800	9096	262	2.88	9174	35	0.38
Beacon	12,813	769	6.00	13,054	328	2.51
Total	80,132	6461	8.06	82,384	1551	1.88
Score band	Bad loans 1 Feb. 2002(set2)			Bad loans 2 Feb. 2002(set2)		
	N	Bad	*N*	Bad	*N*	Bad
0	1840	215	1840	215	1840	215
1–500	231	92	231	92	231	92
501–550	646	189	646	189	646	189
551–600	2106	533	2106	533	2106	533
601–650	5348	1078	5348	1078	5348	1078
651–700	11,624	1641	11,624	1641	11,624	1641
701–750	18,392	1647	18,392	1647	18,392	1647
751–800	20,951	969	20,951	969	20,951	969
≥800	8800	278	8800	278	8800	278
Beacon	17,339	1349	17,339	1349	17,339	1349
Total	87,277	7991	87,277	7991	87,277	7991

data. The bad rates by product line size are less than 10 %. The bad rates decreased with respect to time by both product line and score band, as can be seen from both tables. For example, for accounts less than or equal to 50 M dollars, we can see from the third row of Table 10.2 that the bad rate decreased from 9.46 % and 2.80 % in Feb. 2002 to 8.46 % and 1.85 % in Jan. 2003 respectively.

Results and Discussion

Computation is done in two steps: (1) Score Distribution and (2) Performance Validation. The first step examines the evidence of a score shift. This population consists of the four types of business line of credit (BLOC) products. The second step measures how well models can predict the bad accounts within a 5-month period. This population only contains one type of BLOC account.

Score Distribution

Figure 10.1 depicts the population stability indices values from January 2001 to June 2003. The values of indices for the $50,000 and $100,000 segments show a steady increase with respect time. The score distribution of the data set is becoming more unlike the most current population as time spans. Yet, the indices still remain below the benchmark of 0.25 that would indicate a significant shift in the score population.

The upward trend is due to two factors: time on books of the accounts and credit balance. A book of the account refers to a record in which commercial accounts are recorded. First, as the portfolio ages, more accounts will be assigned lower values (i.e. less risky) by the variable time on books of the accounts, thus contributing to a shift in the overall score. Second, more and more accounts do not have a credit balance as time goes. As a result, more accounts will receive higher scores to indicate riskier behavior.

The shifted score distribution indicates that the population used to develop the model is different from the most recent population. As a result, the weights that had been assigned to each characteristic value might not be the ones most suitable for the current population. Therefore, we have to conduct the following performance validation computation.

Performance

To compare the discriminate power of the SBB scorecard with the credit bureau scorecard model, we depict the Lorenz Curve for both 'Bad 1' and 'Bad 2' accounts in Figs. 10.2 and 10.3. From both Figs. 10.2 and 10.3, we can see that the SBB model still provides an effective means of discriminating the 'good' from 'bad' accounts and that the SBB scorecard captures bad accounts much more quickly than the Beacon score. Based on the 'Bad 1' accounts in January 2003, SBS capture 58 % of bad accounts, and outperforms the Beacon value of 42 %. One of the reason for Beacon model being bad in capturing bad accounts is that the credit risk of one of the owners may not necessarily be indicative of the credit risk of the business. Instead, a Credit Bureau scorecard based on the business may be more suitable.

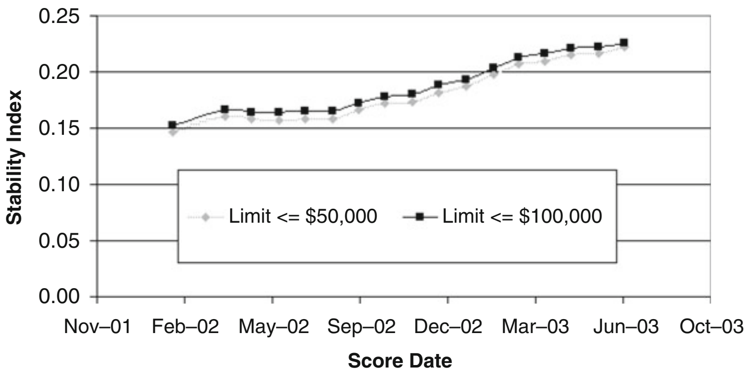

Fig. 10.1 Population stability indices (Jan. 02–June 03)

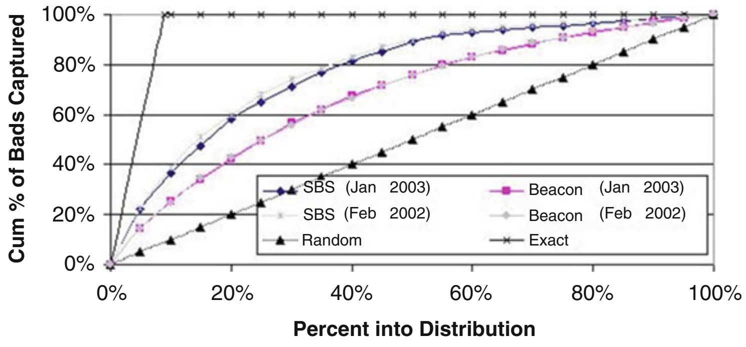

Fig. 10.2 Lorenz curve for 'Bad 1' accounts

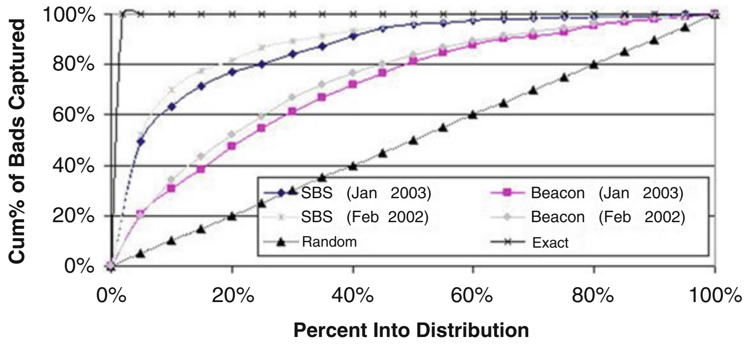

Fig. 10.3 Lorenz curve for 'Bad 2' accounts

Table 10.4 reports various performance statistic values for both 'Bad 1' and 'Bad 2' accounts. Two main patterns are found. First, the Divergence and K-S score values produce consistent results as Lorenz Curve did. For both 'Bad 1' and 'Bad 2', the SBB scorecard performs better than the bureau score in predicting a bad account. Second, SBS based on both bad accounts possibly experience performance deterioration. Table 10.4 shows that all performance statistic based on the January 2003 data are worse than those of the February 2002 period. For example, the 'Bad 1' scorecard generates K-S statistic scores of 78 and 136, for January 2003 and February 2003 respectively. The 'Bad 2' scorecard generates K-S statistic scores of 233 and 394 for both periods.

Table 10.5 gives performance statistic values for both credit lines. i.e., accounts with Credit Limit less than or equal to \$50 M and between \$50 M and 100 M. This table shows a comparison between accounts with a limit of \$50 M and those with limits between \$50 M and 100 M. Two main patterns are found. First, the Small Business Scorecards perform well on both, and outperform the Beacon score on both segments. Second, both scorecards, especially the Small Business Scorecard,

Table 10.4 Performance statistic for both 'Bad 1' and 'Bad 2' accounts

Statistic	SBS (Jan. 2003)	Beacon (Jan. 2003)	SBS (Feb. 2002)	Beacon (Feb. 2002)	SBS (Jan. 2003)	Beacon (Jan. 2003)	SBS (Feb. 2002)	Beacon (Feb. 2002)
# Good	60,542	60,542	61,671	61,671	66,871	66,871	69,312	69,312
Mean good	108.89	738.71	127.3	734.67	137.4	734.28	171.81	729.23
Standard good	172.74	60.18	203.26	63.53	221.22	62.78	284.21	66.66
	'Bad 1' accounts				'Bad 2' accounts			
# Accounts	5567	5567	6427	6427	1196	1196	1879	1879
Mean score	344.9	693.13	439.63	685.79	699.82	678.03	995.65	663.2
Standard deviation	321.53	69.45	387.24	73.27	570.77	75.42	756.34	76.08
Bad rate	8.42 %	8.42 %	9.44 %	9.44 %	1.76 %	1.76 %	2.64 %	2.64 %
Divergence	0.836	0.492	1.02	0.508	1.688	0.657	2.079	0.852
K-S	78	726	136	716	233	726	394	707

Table 10.5 Performance statistics for both credit lines

	Credit line	Limit ≤ \$50 M				Limit \$50–100 M			
	Statistic	SBS (Jan. 2003)	Beacon (Jan. 2003)	SBS (Feb. 2002)	Beacon (Feb. 2002)	SBS (Jan. 2003)	Beacon(Jan. 2003)	SBS(Feb. 2002)	Beacon (Feb. 2002)
Good	# Accounts	47,682	47,682	48,539	48,539	6232	6232	6278	6278
	Mean	116.12	737.77	138.80	733.12	115.13	752.18	125.52	752.64
	Standard	177.34	59.12	213.62	62.52	161.93	54.61	174.07	55.86
Bad	# Accounts	4393	4393	5226	5226	545	545	637	637
	Mean score	347.40	695.10	461.06	686.03	345.82	715.80	398.05	711.95
	Standard deviation	314.69	65.68	391.94	71.87	285.01	68.35	310.59	62.28
Performance	Bad rate	8.44 %	8.44 %	9.72 %	9.72 %	8.04 %	8.04 %	9.21 %	9.21 %
	Divergence	0.820	0.466	1.042	0.489	0.991	0.346	1.172	0.473
	K-S	78	726	136	717	125	735	162	742

perform better on 'Bad 2' accounts. The main reason is that 'Bad 2' definition specifies a more severe degree of delinquency and the difference between the good and bad accounts is more distinct.

Conclusions

Balanced scorecard analysis provides a means to measure multiple strategic perspectives. The basic principle is to select four diverse areas of strategic importance, and within each, to identify concrete measures that managers can use to gauge organizational performance on multiple scales. This allows consideration of multiple perspectives or stakeholders. Examples given included supply chain risk analysis, and policy analysis of natural gas vehicle adoption. This chapter focused on the example of a small bank credit situation. Computation results indicate there is evidence of a shifting score distribution utilized by the scorecard. However, the scorecard still provides an effective means to predict 'bad' accounts.

Balanced scorecards have been widely applied in general, but not specifically to enterprise risk management. This chapter demonstrates how the balanced scorecard can be applied to evaluate the risk management posture of a particular organization. The demonstration specifically is for a bank, but other organizations could measure appropriate risk elements for their circumstances. Balanced scorecards offer the flexibility to include any type of measure key to production planning and operations of any type of organization.

Notes

1. Kaplan, R.S. and Norton, D.P. (2006). *Alignment: Using the Balanced Scorecard to Create Corporate Synergies*. Cambridge, MA: Harvard Business School Press Books.
2. Olhager, J. and Wikner, J. (2000), Production Planning and Control Tools. *Production Planning and Control* 11:3, 210–222.
3. Al-Mashari, M., Al-Mudimigh, A. and Zairi, M. (2003). Enterprise resource planning: A taxonomy of critical factors. *European Journal of Operational Research*, 146:2, 352–364.
4. Alquier, A.M.B. and Tignol, M.H.L. (2006). Risk management in small- and medium-sized enterprises. *Production Planning & Control,* 17, 273–282.
5. Kaplan and Norton (2006), op cit.
6. Elbannan, M.A. and Elbannan, M.A. (2015). Economic consequences of bank disclosure in the financial statements before and during the financial crisis: Evidence from Egypt. *Journal of Accounting, Auditing & Finance* 30(2), 181–217.
7. Schaefer, A. Cassidy, M., Marshall, K. and Rossi, J. (2006). Internal audits and executive education: A holy alliance to reduce theft and misreporting. *Employee Relations Law Journal*, *32*(1), 61–84.
8. Beasley, M. Chen, A., Nunez, K. and Wright, L. (2006). Working hand in hand: Balanced scorecards and enterprise risk management, *Strategic Finance* 87:9, 49–55.
9. Campbell, M. Adams, G.W., Campbell, D.R. and Rose, M.R. (2006). Internal audit can deliver more value, *Financial Executive* 22:1, 44–47.

10. Sugarman, P. and Kakabadse, N. (2008). A model of mental health governance, *The International Journal of Clinical Leadership* 16, 17–26.
11. Janssen, A., Lienin, S.F., Gassmann, F. and Wokaun, A. (2006). Model aided policy development for the market penetration of natural gas vehicles in Switzerland, *Transportation Research Part A* 40, 316–333.
12. Wu, D.D. and Olson, D.L. (2009). Enterprise risk management: Small business scorecard analysis. *Production Planning & Control* 20(4), 362–369.