Journal of Manufacturing Systems 48 (2018) 166–179

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

A statistical system management method to tackle data uncertainty when using key performance indicators of the balanced scorecard

R. Sanchez-Marquez[a,*], J.M. Albarracin Guillem[b], E. Vicens-Salort[c], J. Jabaloyes Vivas[d]

[a] Departamento de Organización de Empresas, Universitat Politècnica de València, Camino de Vera, s/n., 46021, Valencia, Spain
[b] Departamento de Organización de Empresas, Universitat Politècnica de València, Spain
[c] Production Management and Engineering Research Centre, Universitat Politècnica de València, Camino de Vera, s/n., 46021, Valencia, Spain
[d] Department of Statistics, Operational Investigation and Quality, Universitat Politecnica de Valencia, Spain

A R T I C L E I N F O

Keywords:
Key performance indicators (KPIs)
Operating system (OS)
Significant trend analysis (STA)
Significant shift analysis (SSA)
Manufacturing

A B S T R A C T

This work is focused on the development of a graphical method using statistical non-parametric tests for randomness and parametric tests to detect significant trends and shifts in key performance indicators from balanced scorecards. It provides managers and executives with a tool to determine if processes are improving or decaying. The method tackles the hitherto unresolved problem of data uncertainty due to sample size for key performance indicators on scorecards. The method has been developed and applied in a multinational manufacturing company using scorecard data from two complete years as a case study approach to test validity and effectiveness.

1. Introduction

Kaplan and Norton's balanced scorecard theory [1–3] has become one of the most common methods for managing performance and especially in large organisations [4]. Some of the theory's limitations and problems are addressed in various studies [5–7].

The use of the balanced scorecard (BSC) as a performance management system (PMS) and its main objective (which is to translate strategy into specific actions) has been studied in many research works [2–4,7–9]. The validity and effectiveness of its scientific use, combined with analytical and other systemic methods, has been confirmed in several investigations [10–14]. These research works are focused on choosing the most important KPIs and proving and quantifying the impact of company strategies and actions.

Several problems and limitations have also been raised by these authors including: sample size (which implies a long period to take enough data points); uncertainty in information; and a high level of expertise needed to apply the methods [10,12].

KPIs from the scorecard indicate performance in each period. Typically, they are monitored on a monthly basis. The objective is to show the performance of the processes that are behind each KPI from different operating systems (OS) or dimensions [13]. Random changes (shifts and drifts) are normal because monthly numbers are based on samples that serve to estimate the KPIs. The one-month cut off is artificial in the sense that the same indicator could be estimated on a weekly or bi-monthly basis. Indeed, it is common to have different periods for different KPIs: weekly, monthly, quarterly, and so on. The same process would show different numbers depending on the period considered (sample). Theoretically, in a continuous variable (KPIs are either proportions or rates) the probability of having exactly the same number is zero. Within KPI estimation, the larger the sample size – the smaller the data uncertainty. The estimation of a confidence interval (CI) and rules for the detection of trends are necessary to distinguish between natural random variation due to sample size; and systemic significant changes made on purpose for process improvements or due to unexpected decay processes.

The traditional way to analyse changes on scorecard KPIs is confusing. Data uncertainty due to sample size drives to the wrong

Abbreviations: ABS, absenteeism; BSC, balanced scorecard; Ci, confidence interval; Cl, confidence level; CLT, central limit theorem; CPU, cost per unit; D/1000, defects per thousand; DOS, delivery operating system; FA, factor analysis; FC, fixed cost; FTT, first time through; KPI, key performance indicator; L&OH CPU, labour and overhead cost per unit; LB, lower bound; LTCR, lost time case rate; OEE, overall equipment effectiveness; OS, operating system; PCA, principal component analysis; PMS, performance management system; POS, people operating system; PTS, production to schedule; RPT, repairs per thousand; SPC, statistical process control; SQDCPME, safety, quality, delivery, cost, people, maintenance, environment; SSA, significant shift analysis; SSMM, statistical system management method; STA, significant trend analysis; TTP, throughput to potential; UB, upper bound

* Corresponding author.
E-mail addresses: rsanch18@ford.com (R. Sanchez-Marquez), jmalbarr@omp.upv.es (J.M. Albarracin Guillem), evicens@cigip.upv.es (E. Vicens-Salort), jabaloye@eio.upv.es (J. Jabaloyes Vivas).

https://doi.org/10.1016/j.jmsy.2018.07.010
Received 9 March 2018; Received in revised form 25 July 2018; Accepted 27 July 2018

conclusions, and therefore, to wrong decisions or inaction. Current practices, based on a deterministic approach, needs to be replaced by methods that tackle data uncertainty due to sample size.

The only attempt within the current literature to tackle the problems of data uncertainty within the balanced scorecard has been made by Breyfogle [15]. He proposes applying statistical process control (SPC) methods from control charts directly on BSC KPIs. This approach, which we also tested, did not work properly for these reasons:

1) Normality assumption is needed for SPC since normal approximation methods without adjusted point estimates are used for CIs. This cannot be confirmed for most KPIs.
2) For KPIs where normality was confirmed, the method implies changing the sampling approach from 100% of units produced per month to one based on subgroups. This implies drastically reducing sample size – which diminishes the power of tests and increases data uncertainty and the number of calculations needed. Average and range/sigma are needed for each subgroup. Such changes make the method more complicated to implement and less precise.
3) In SPC, CIs are estimated using a confidence level (1-α) of 99.73% (± 3σ) because they are estimated from a stable process and the purpose is continuing within those limits. The main purpose of BSC is to detect KPIs and/or dimensional improvements in the achievement of corporate goals and objectives. Therefore, confidence levels recommended for hypothesis testing (95% or 99%) are better for application on BSC KPIs.
4) The autocorrelation effect is usually present in time series. SPC methods do not take into account autocorrelation to avoid the false detection of significant trends.

Within this paper, we present a proposal for a statistical system management method (SSMM). We developed the idea as suggested by Breyfogle [15] by tackling and improving its problems and limitations. We used as a starting point a group of main KPIs that were selected applying the methods developed in other research works [10–13]. A complexity reduction of the BSC was vital since at the beginning of the research it was composed of almost 90 KPIs.

This work dealt with the development of a methodology based on tests for significant shift analysis (SSA) and significant trend analysis (STA) using the application of the most appropriate parametric and non-parametric statistical test for randomness (hypothesis test) for each KPI. This method tackled uncertainty due to sample size. Uncertainty due to data integrity was considered negligible for all processes since the company where the method was developed and tested applied techniques for measurement system analysis (such as Gage R&R and calibration). The company was ISO 9001certified. Uncertainty due to data integrity was not within the scope of this research work.

Within the results analysis and discussion section, we checked the effectiveness of each test by applying it to the real scorecard of a manufacturing company in a case study approach. We worked on this research project in the context of a collaboration agreement between the Universitat Politècnica de València and the company (a multinational global leader in the automotive industry). The research work was proposed by the company as part of their strategic initiative for improving management methods. The method was implemented for the balanced scorecard of the Spanish subsidiary company and was included in future strategies to be implemented globally.

This company uses the approach of seven OS/dimensions SQDCPME [[13] [16],] for the BSC.

The purpose of this paper is not to develop new statistical methods. It is to develop a procedure based on a graphical method for managing data uncertainty due to sample size within the BSC. It was based on the most appropriate statistical methods to estimate CIs for each KPI and using the methods to design a graphical hypothesis test to detect significant shifts. Additionally, we also designed a graphical hypothesis test for significant trend detection based on the best available methods for non-parametric tests – including a correction for the autocorrelation effect of the time series.

2. Literature review

The review of the current literature was focused on three objectives. The first objective was to assess the appropriateness of the use of statistical tools, which is in essence, a qualitative analysis. The second and third objectives were to review and select the most appropriate test for each KPI (a mix of qualitative and a quantitative analysis).

The graphical method we are proposing and developing in this paper aims for two types of change detection. Firstly, process drift by means of the identification of significant trends on KPIs, or significant trend analysis (STA); and secondly, process shifts by means of the identification of significant changes from the previous month, or significant shift analysis (SSA).

In a similar way to SPC control charts, trends will be detected using non-parametric statistical tests for randomness. Shifts from month to month have to be detected using the parametric test that best fits each KPI. However, due to the reasons mentioned in the introduction section, we cannot use the same techniques as SPC.

The main KPIs taken from BSC can be classified into two groups. The first group is composed of metrics defined by binomial proportions: a delivery operating system (DOS); *PTS* (production to schedule); a people operating system (POS); and absenteeism, etc. The second group of metrics is composed of those defined as rates. These include: *LTCR* (lost time case rate) for safety; warranty repairs (also counted per thousand units sold because frequency is low to be expressed per unit for quality); and internal repairs per thousand units built for offline repairs. Both metrics reflect the number of defects per unit found in the field or in production.

Table 1 below, summarises the KPIs selected in the case study from

Table 1
Main KPIs selected from the balanced scorecard.

Operating System	Acronym/Abbreviation	Name	Description	Units
Safety	*LTCR*	Lost time case rate	Number of accidents every 200,000 working hours	Accidents/ 200,000 h
Quality	*RPT 3MIS*	Warranties *RPT* @ 3 *MIS*	Number of repairs at 3 months in service every 1000 units sold (costumer claims)	Repairs/ 1000 units
Quality	*Offline*	Offline repairs	Internal repairs made on the units outside the production normal flow (off-line)	Repairs/ 1000 units
Delivery	*PTS*	Production to schedule	Proportion of units produced according to daily production schedule	%
Cost	*L&OH CPU*	Labour and other overhead cost per unit	Labour costs and other related costs such as industrial supplies per unit produced	$/unit
People	*ABS*	Absenteeism	Proportion of people that do not attend work on a daily basis due to unexpected reasons (e.g. illness)	%
Maintenance	*TTP*	Throughput to potential	Proportion of units produced over the demand-adjusted capacity expressed in units	%