Full Length Research Paper

MS excel functions as supply chain fraud detector

Tej Narain Prasad Varma[1*] and DA Khan[2]

Department of Computer Applications, NIT, Jamshedpur, India.

Information technology (IT) is playing a vital role in increasing the productivity, profitability of businesses and optimizing decisions at each stage of Supply Chain Networks (SCN). Supply chain manager must often trust data for decision making even reported from vendors/suppliers. These types of data are vulnerable to manipulation and creating an opportunity for 'supplier opportunism'. Use of MS excel functions is a simple solution. The purpose of this empirical study is to find the intentionally manipulated data with help of excel functions. This finding can allow supply chain managers to segregate suspect data from decision-making until they can be validated and thus mitigate supplier opportunism.

Key words: Information technology, MS excel function, supply chain.

INTRODUCTION

In a digital age, internet has become the necessity of life that generates e-risks by fraudsters, though data interception, data interference, system interference or illegal access by e-mail spoofing or forgery, phishing, email spam, Denial of service attacks, unauthorized access physically or virtually to computer/computer system/computer networks, web jacking physically damaging the computer system etc are exponentially growing the addition cost to the organization/government to manage the e-risk in their supply chain networks. Fraud is a deceit, trickery, sharp practice or breach of confidence, perpetrated for profit or to gain some unfair or dishonest advantage (dictionaty.com). Broadly fraud can be categorized into the three categories: Asset misappropriations—Involving the theft or misuse of an organization's assets; Corruption—When fraudsters wrongfully use their influence in a business transaction to procure some benefit for themselves or another person,

contrary to their duty to their employer or the rights of another; Fraudulent statements—Involving the falsification of an organization's financial statements .Within the above three global categories, ACFE (2013) identifies more than 70 areas of fraud. Organizational fraud can be classified into 15 categories: Bribery/illegal gratuities/ economic extortion, Conflicts of interest, Fictitious revenues/timing differences, Understated liabilities and expenses, Overstated assets/valuation, improper disclosures, Non-financial fraudulent statements Cash larceny, Skimming ,Inventory misuse/larceny, Billing schemes, Payroll schemes, Expense reimbursement schemes, Check tampering and Register disbursements. Supply chain data are soft targets for fraud by asset misappropriation, bid rigging, phantom bids, nepotism, substitution, false count, counterfeiting, creating fictitious accounting entities e.g., ghost employee, fake vendor, fake customer or vendor payments, falsified hours etc.

*Corresponding author. E-mail: tpverm@yahoo.com.

It is very challenging for detective agencies to find frauds that occur in virtual supply chain environment. In this era, the scenario has undergone tremendous change because bytes are replacing bullets in the crime world. The Computer Assisted Audit Tools (CAATs) software and digital tools are absolutely essentials for these agencies but these are costly. Spreadsheets are one of the most popular and ubiquitous software packages on the planet. Every day, millions of business people use spreadsheet programs to build models of the decision problems they face as a regular part of their work activities (Ragsdale, 2007), forecasting with Excel (Radovilsky and Eyck, 2000), An Excel Based Case Using Financial Statement Analysis to Detect Fraud (Ragan, 2008) etc. This paper identifies the different MS Excel Functions and investigates its role and applications in managing supply chain frauds. The details of the research methodology are presented in Section 3. A brief review of the Excel function is presented in Sub-Section 4.1 and 4.2 related to data cleaning for analysis. Further onward section 4.3 of section 4, excel function with its use in digital analysis is discussed which was also tested as fraud detector on supply chain data of an organization. In discussion section the different types of supply chain frauds were reviewed for detection. Finally, result and conclusions are presented.

Microsoft excel

Microsoft excel is a spreadsheet which helps us to organize data in rows and columns of cells and it is simpler than most CAAT tools . It is also highly flexible, with huge list of functions, possible to install Add-Ins with advanced features, powerful Data Import feature, and lower cost for installation. Microsoft Excel has many powerful features and by using this can easily detect and prevent fraudulent activity; it has some limitations: it cannot log or document the audit work done, involves complex procedures to do detailed analysis, is prone to errors/tampering as data is open, risk of hidden rows and columns, takes much longer to process large data and data size limitation of processing only one million rows or records of data (Excel, 2007). Data lying in any of the formats: Text Files (*.TXT), Comma Delimited Files (*.CSV), Database Files (*.DBF), Extendible Markup Language Data Files (*.XML), Microsoft Access Database (*.MDB), Lotus 123 Databases (*.WK_) can easily be opened in Excel. It also supports Import of External Data through ODBC. Using SQL Data Queries, perhaps any data source can be accessed. This could range from SQL Server, Oracle to IBM DB2 database.

Excel function

Functions are formulas that Excel has predefined. An Excel function is a preset formula that calculates a specific result based on the criteria/variables/arguments; all functions start with the equal sign followed by the function's name and criteria/variables/arguments. It makes simple but cumbersome formulas easier to use, enables one to include complex mathematical expressions in worksheets that otherwise would be difficult or impossible to construct using simple arithmetic operators and enables one to include data in applications that could not be accessed otherwise. However, Excel's logical functions are designed to create decision-making formulas (McFedries, 2007, 2010).

RESEARCH METHODOLOGY

The study was conducted to review different Microsoft excel functions, which were part of Excel version 2007, 2010 and 2013. These functions were reviewed and sorted on the basis of their utilization in digital analysis. Further, these sorted functions were tested on an organizational data related to invoice data of vendor, sales data of customer as cleaning of data, counting, stratifications, ageing, gap analysis, duplicate check, etc. It can be replicated similar type of detection as duplicate check may be used for checking duplicate vendor, duplicate invoice etc. Further, in discussion section, the different types of supply chain frauds were reviewed and use of excel function for detection is also discussed. Data were collected primarily through books on Excel of different versions and from some e-journals search engines available in intranet or organisational library that are in the areas of digital analysis, excel and excel function, fraud etc.

MS excel functions as fraud detector

The commonly used Audit softwares are ACL, ActiveData For Excel, IDEA, TopCAATs etc. But Microsoft Excel is an ever-present tool and easily use for data analysis because spreadsheets are easy to navigate and flexible enough. Excel allows users to calculate a specific result on the basis of the user criteria, which start with the equal sign followed by the function's name; and criteria or variables or arguments as Sort, Subtotal, Filter, and Merge data etc. This is done by using inbuilt functions and it performs statistical analysis also. There are a total of 455 Excel Functions in Excel 2013; however 299 functions were in 2001 version (Excel 5.0) as shown in Appendix I and categories wise shown in Appendix II. These functions are separated into compatibility, cubes, databases, date and times, engineering, financials, information, logical, lookup and references, math and trigonometry, statistical, texts, user defined add-ins and webs. Hence, Microsoft Excel has become an industry standard in managing and analyzing organizational data. Some Excel functions as IF, combination of IF, AND, SUM, OR, VLOOKUP are very powerful tools for auditing

or investigating agencies for detecting fraud from supply chain networks.

Excel data cleaning functions

During investigation or auditing we receive or retrieve data in different formats such as Text, Comma Separated Value (CSV) or Web Page formats. For analyzing these retrieved data in excel, this is a requirement to clean it first to match the criteria that we specify. There is so many features in excel to perform the cleaning activities. Some of the following excel functions for cleaning data are: CLEAN function (syntax CLEAN (text)) removes all non-printable characters from text. This function can be used on text imported from other applications that contain characters that may not print with certain operating system. For example, we can remove some low-level computer code that is frequently at the beginning and end of data files and cannot be printed. TRIM function (syntax is TRIM (text)) removes all spaces from text except for single spaces between words. TEXT function converts a value to text in a specific number format. Its syntax is TEXT (value, format_text), where value is a numeric value, a formula that evaluates a numeric value or a reference to a cell containing a numeric value; and format_text is a number format in text form from the Category box on the Number tab in the Format Cells dialog box. CONCATENATE function joins several text strings into one text string. Its syntax is CONCATENATE (text1, text2,...). The "&" operator can be used instead of this function to join text items.

LEFT Function (syntax LEFT (text, num_chars)) returns the first character or characters in a text string, based on the number of characters specified. Similarly RIGHT returns the last character or characters in a text string, based on the number of characters specified. MID returns a specific number of characters from a text string, starting at the position we specify, based on the number of characters specified. FIND function finds one text string (find_text) within another text string (within_text), and returns the number of the starting position of find_text, from the first character of within_text. SEARCH function returns the number of the character at which a specific character or text string is first found, beginning with start_num. Use SEARCH to determine the location of a character or text string within another text string so that we can use the MID or REPLACE functions to change the text. But unlike SEARCH, FIND is case sensitive and does not allow wildcard characters. REPLACE function replaces part of a text string, based on the number of characters we specify, with a different text string. SUBSTITUTE function substitutes new_text for old_text in a text string. This function is used when we want to replace specific text in a text string; but REPLACE function is used when we want to replace any text that occurs in a specific location in a text string. LOWER

function converts all uppercase letters in a text string to lowercase and similarly UPPER function converts text to uppercase. PROPER function capitalizes the first letter in a text string and any other letters in text that follow any character other than a letter.

It converts all other letters to lowercase letters. FIXED function rounds a number to the specified number of decimals, formats the number in decimal format using a period and commas, and returns the result as text. Its syntax is FIXED (number, decimals, no_commas), where number is the number we want to round and convert to text, decimals is the number of digits to the right of the decimal point and no_commas is a logical value that, if TRUE, prevents FIXED from including commas in the returned text. LEN function returns the number of characters in a text string. VALUE function converts a text string that represents a number to a number. CODE function returns a numeric code for the first character in a text string.

The returned code corresponds to the character set used by our computer. CHAR function returns the character specified by a number. Use of this function is to translate code page numbers we might get from files on other types of computers into characters. CELL function returns information about the formatting, location, or contents of the upper-left cell in a reference.
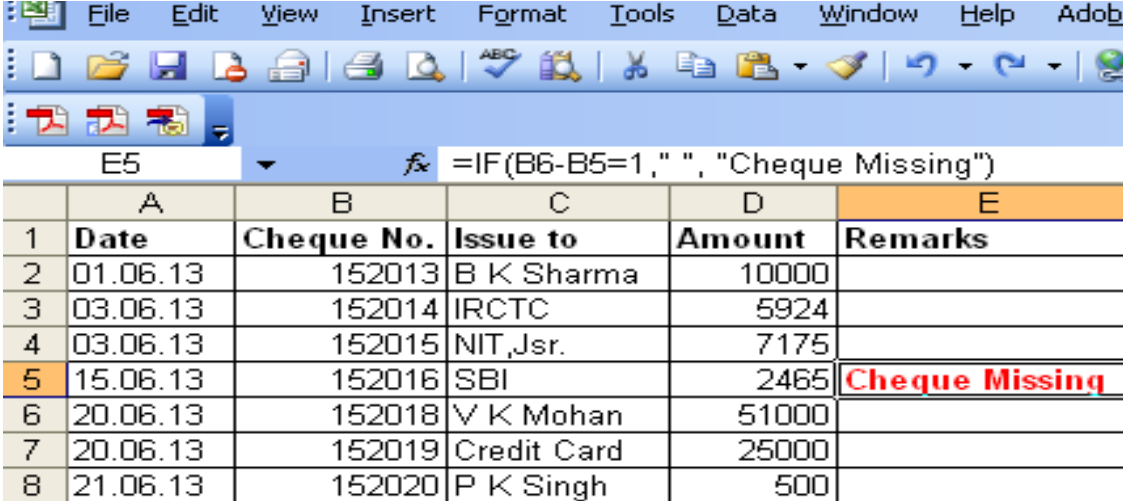DOLLAR function converts a number to text format and applies a currency symbol. RTD function retrieves real-time data from a program that supports COM automation. TRIMMEAN function returns the mean of the interior of a data set. This function can be used for excluding outlying data from analysis.

Rounding and formatting function

ROUND function rounds a number to a specified number of digits and MROUND returns a number rounded to the desired multiple. TRUNC function truncates a number to an integer by removing the fractional part of the number. CEILING function returns number rounded up, away from zero, to the nearest multiple of significance and similarly FLOOR rounds number down, toward zero, to the nearest multiple of significance. INT rounds a number down to the nearest integer, ODD returns number rounded up to the nearest odd integer, EVEN returns number rounded up to the nearest even integer and ROUNDDOWN rounds a number down, toward zero.
The evaluations of financial information made by a study of plausible relationships among both financial and non-financial data to assess whether account balances appear reasonable (AICPA, SAS 56) for analytical Tests like horizontal analysis (increase or decrease over two or more periods), vertical analysis, ratio, etc by simple excel functions.

The following functions are more useful to detect frauds in supply chain environment.

Figure 1. Gap analysis of missing document.

IF function

The most common and powerful of the logical functions in Excel is the IF function. The syntax of IF function is '=IF (logical_test,value_if_true,value_if_false)'.The IF function is used for ageing analysis, gap detection, duplicate records finding, locating multiple records, extracting records meeting certain criteria etc. This is very useful function of decision making. IF function had been used for finding sequential missing of data as shown in Figure 1.

Gap analysis by using IF function

This is a tool which is used for identifying gaps of missing within a specified field in a file for sorted and indexed data. These gaps analysis may be applied for manipulated or intentionally deleted entry from serial sequence control documents, products identification, certification, human resource data etc. Excel makes it possible to identify if there are any gaps in any column of data which is expected to have a sequential numbering. It can be accomplished in a very simple manner. We can sort the data serially on the said column and then in a new column, calculate the difference of the value from the value in the above row. The calculated values should be 1 if there is no gap. We are using IF function. Given screen sort for missing cheque of an individual data, the syntax for E5 field is '=IF (B6-B5=1," ", "Cheque Missing")'

Nested IF function

This is IF within IF function. It is used when evaluating either the *value_if_true or value_if_false* arguments.

Use of nested IF in stratifications, count and ageing

Stratifications, counting and ageing for SCM data provide a useful view into the largest, smallest, and average transactions within specified intervals/group. It is also an important tool for auditors and detecting agencies for analytical test of data (Figure 2).

IF function with AND / OR function

If simultaneous confirmation for logical actions is required then we may use AND or OR with the IF function. It is often necessary to perform an action if and only if two conditions are true. In Excel, And conditions are handled, appropriately enough, by the AND() logical function: AND(logical1 [, logical2 ,...]). Similar to an And condition is the situation when you need to take an action if one thing or another is true. Or conditions are handled in Excel by the OR() function: OR(logical1 [, logical2 ,...]).

Excel function as duplicate checker

This is converse of gap analysis where the serial control sequence should not be repeated. It is used for detecting duplicate data within a specified field in a file, e.g. duplicate vendors, duplicate invoices and other duplicate entries which lead to fraudulent transaction in the supply chain networks. It can be also applicable where same-same-different checks are required by auditing or detecting agencies as cluster of vendor code and their invoice numbers for excess or double payment to fraudulent vendor. Here excess weight done in the system against tare memo no. 218 against a sell data of an organization and its syntax appears in the given screen shot (Figure 3).
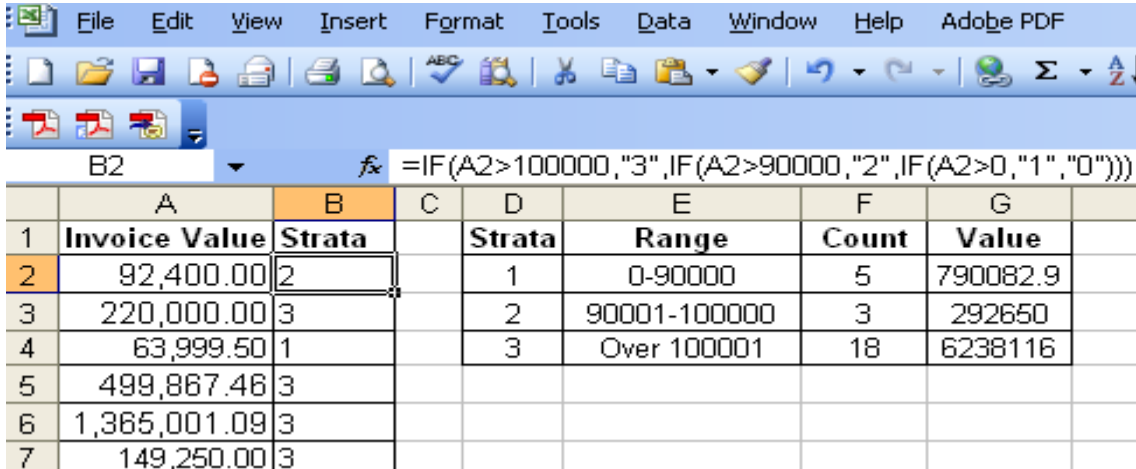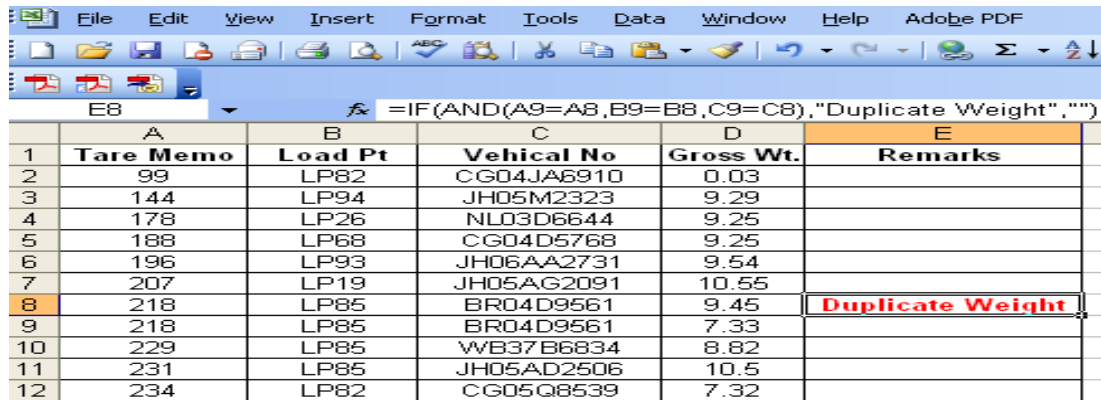
Figure 2. Stratifications and count.

Figure 3. Duplicate entry checking.

Count IF function

The Count IF function is used to analyze the master data (material, service, vendor, customer, employee etc.) of any organizations with the user criteria. It helps the investigator in filtering, sorting, duplicating verification and applying Benford's analysis etc. The syntax of Count IF function is '=COUNTIF (range, criteria)'.
The screen shot (Figure 4) shows the number of time the tare memo appears in column A , which indicates that there is a violation of serial control mechanism. This can be appliied to check the duplicate vendor though different entities like telephone number, tax code, address etc.
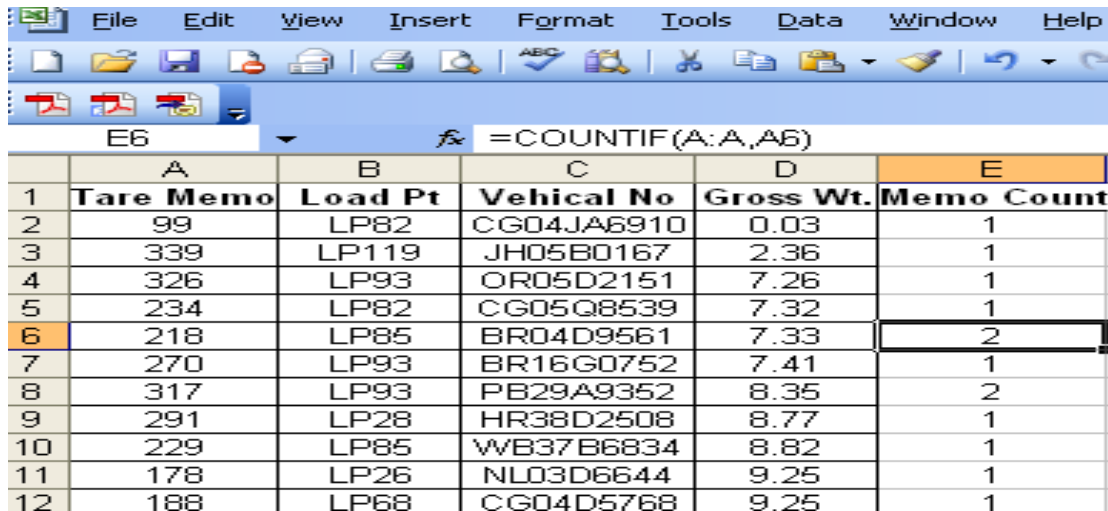
SUMIF function

The summation of the cells value of specified by a given/user defined criteria can be performed with the application of SUMIF function. This function works as Pivot Table command but this function is useful where real time data are analyzed. It may be used for strata calculation.   Its   syntax   is   "=SUMIF (range,criteria,sum_range)".
   In Figure 5, it appears that vehicles No. enter in cell C6 loaded more than one time with weight appearing in cell value E6.
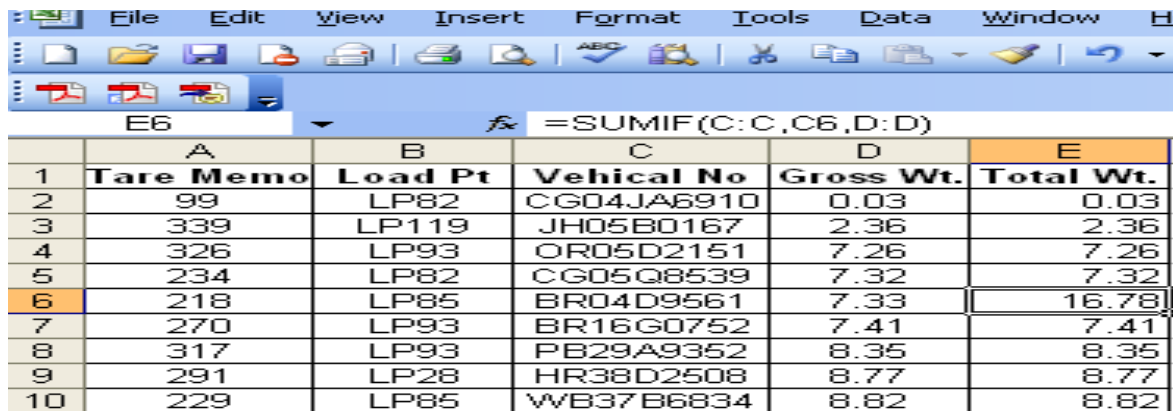
VLOOKUP Function

This function is useful for auditor in combining specified fields from two different files into a single file using key fields. It creates relational databases on key fields. There are several applications of this function in auditing or investigation which depends upon the circumstances and subject   matters.   The   syntax   is   "=VLOOKUP (lookup_value,table_array,col_index_num,range_lookup)" , where Lookup_value is the value to be found in the first column of the array. Lookup_value can be a value, a reference, or a text string; Table_array   is the table of information in which data is looked up. Use a reference to a range or a range name, such as Database or List, Col_index_num   is the column number in table_array

Figure 4. Counting of entry.

Figure 5. Use of SUMIF.

from which the matching value must be returned and Range_lookup is a logical value that specifies whether we want VLOOKUP to find an exact match or an approximate match.

Join / Relate for master data

This combines specified fields from two different files into a single file using key fields. This function is used to create relational databases on key fields. For example, the vendor master file could be related to the invoice file to obtain address information for cheque clearing bank account no. against each invoice.

LOOKUP function

LOOKUP functions are functions which lookup the value of a selected cell in another data table and return the corresponding detail of the matching cell. Using LOOKUP

functions we can create a MASTERTRANSACTION, or a PARENT-CHILD relationship between different data tables. Lookup functions can be used to check the data integrity of tables and also to link data from multiple tables. This function returns a value either from a one-row or one-column range or from an array. The LOOKUP function has two syntax forms: vector and array. The vector form of LOOKUP looks in a one-row or one-column range (known as a vector) for a value and returns a value from the same position in a second one-row or one-column range. The array form of LOOKUP looks in the first row or column of an array for the specified value and returns a value from the same position in the last row or column of the array. The array form of LOOKUP looks in the first row or column of an array for the specified value and returns a value from the same position in the last row or column of the array. Use this form of LOOKUP when the values we want to match are in the first row or column of the array. Use the other form of LOOKUP when we want to specify the location of the column or

row. The array form of LOOKUP is very similar to the HLOOKUP and VLOOKUP functions. The difference is that HLOOKUP searches for lookup_value in the first row, VLOOKUP searches in the first column, and LOOKUP searches according to the dimensions of array.

RAND and RANDBETWEEN Function

Excel provides two tools for generating random numbers. The RAND function (syntax "RAND()") in Microsoft Excel allows to generate random numbers from the uniform distribution. It is a volatile function, which means it will be recalculated any time the enter key is pressed, so the random number constantly changes. RANDBETWEEN returns a random number between the numbers as specified.

Cross Tabulate

Cross Tabulate analyzes character fields by setting them in rows and columns. By cross tabulating character fields, we can produce various summaries, explore areas of interest, and accumulate numeric fields. Excel effectuates cross-tabulation through its Pivot Tables .The pivot table quickly summaries or analyses large data as subtotaling, aggregating, categorizing, creating custom calculation and formula. It also helps in filtering, sorting and grouping the data. Because a Pivot Table report is interactive, we can move rows to columns or vice-versa. To apply the Pivot Table, we firstly select the icon and navigation through interactive dialogue box. By selecting appropriate range of data and output table area, it is easily crosstabs as row and column labels. It is a data consolidate approach that sums automatically, can be used as pivot, drillable, table formatting etc. for detecting frauds.

Data analysis tools

The Data Analysis add-in has much more than just a Correlation tool. It includes a tool that returns descriptive statistics for single variable, tools for several inferential tests (Carlberg, 2010). By help of univariate Statistics (statistics relating to a single variable) we can find out the mean, median, mode, standard deviation, skewness, kurtosis and various other statistical data relating to the variable. This gives a general idea about the behavior pattern of data but also forms a support base for conducting further statistical analysis of the data for fraud detection.

DISCUSSION

There are possibilities of different type of fraud risk in supply chain network as bid rigging, phantom bids,

nepotism, substitution, false count, counterfeiting, creating fictitious accounting entities e.g., ghost employee, fake vendor, fake customer or vendor payments, falsified hours etc. We can easily detect and prevent fraudulent activities by help of above discussed excel function. There is some following application of excel sheet functions which were discussed above in overcoming the fraud in supply chain.

1. Conflict of Interest - Officials involved in Supply Chain Management have to act their duties in organizational interest. If they perform duties to gain any benefits to their family member or friends is the part of conflict of interest. For detection this matching algorithms are suitable as employee – vendor key field as telephone no, e-mail address, social ID no, bank account number by using IF, MATCH, VLOOKUP etc. functions. As shown in Figure 3 can be replicate for detecting this fraud.

2. Fake customer or vendor Payment – Duplicate payments are far more common than most organisations realize or are prepared to admit. *Core Algorithms* is a logic that identifies all duplicate payments by using four fields of vendor payment i.e; vendor code, invoice number, invoice date and invoice amount with exact (E),Similar (S) and different (D) combination matching on these fields. The possible combinations are EEEE (means exact matching of vendor code, invoice no., date and amount), DEEE, ESEE, EESE, EEES, ESES, ESSE, EESS and DESE. We can use IF, combination of IF AND function for getting the core algorithms combinations. The type of fraud is detected by using this core algorithms common name and address of employee for payment. As shown in figure 3 can be replicate for detecting this fraud.

3. Duplicate vendor – This is the occurrence of multiple vendor code in name of any vendor's proprietor. Generally, auditors or detecting agencies are using the address, tax id, and bank number or vendor sensitive data. For this IF with AND function and the condition formatting is useful for duplicate check. As shown in Figure 3, it can be replicated for detecting this fraud.

4. Ghost Employee – It is someone on the payroll who does not actually work for a victim organization. It can be detect by analysis of payments being made to employees that are not reconciling to other independent employee data sources.

5. Piggyback Fraud – This is the malicious entry with valid of person or data or material. This can be detected by comparing or matching data by using IF function.

6. Missing Vendor or Customer data – IF function may be used for identify the blank fields of vendor or customer data. As shown in Figure 2, it can be replicated for detecting this fraud.

On application of excel function on the sample data (although the amount of individual data is proprietary) of an organization, has been selected from procurement cycle of Supply Chain Network, vendor master, employee master, etc. for verification of above mention supply chain fraud because vendor and employee are essential components of supply chains. As discussed above during our experiment, duplicate invoices had been detected. Further, duplicate check was replicated on vendor master data extracted in excel sheet from SAP system with vendor code and its telephone nos. The vendors with common telephone numbers were detected in the system and they were eliminated as further action. The vendor and employee with common mobile numbers or address were also detected, which was the part of conflict of interest. During the course of work, a concern was received that there was profiteering by several vendors in this organization. Prima facie verification of records reflected that the amounts billed by the contractors were logical. At this point the combined function as LEFT for finding significant digit (FSD), SUBTOTAL,COUNTIF for frequency of FSD has been implemented during detecting supply chain fraud by using Benford's distribution for same organization. As detected most payments were made between Rs.90000 to Rs.99999 to avoid the higher approval authorities limit, i.e. one lakh, which was the generation of procurement fraud due to splitting of purchase orders, and the repeat orders were awarded to the vendors to manipulate payment with ulterior motives (Varma and Khan, 2013). On further analysis, it was observed that the these fraudulent activities were made due to incomplete or vague job specification, selection of vendor without proper capability assessment, and wrong inputs/ incomplete data in negotiation sheet to highlight capability of vendor, etc., which lead to a major investigation after which strong nexus between officers and vendors was found .

Conclusion

Microsoft Excel can be used to develop some management decision making supporting tools for the sake of easy use and low cost of ownership. We illustrated application in supply chain for auditing and investigating. Technology is always a double-edged sword. Society that is dependent more and more on technology, cyber crimes are bound to increase because bytes are replacing bullets in the crime world. History is the witness that no legislation has succeeded in totally eliminating crime. The advanced quantitative modeling techniques (such as multiple regression, etc.) are involved; Excel add in called Data Analysis should be activated and use for quick and advance analysis with big data.

Conflict of Interests

The authors have not declared any conflict of interests.

REFERENCE

ACFE (2013). "Report to the Nation on Occupational Fraud and Abuse" The Association of Certified Fraud Examiners, http://www.acfe.com

Carlberg C (2010). Statistical Analysis: Microsoft Excel', Indianapolis, IN: Que Publishing

McFedries P (2007). Formulas and functions with Microsoft Office Excel 2007. Indianapolis, IN: Que Publishing

McFedries P (2010). Formulas and functions with Microsoft Office Excel 2010. Indianapolis, IN: Que Publishing

Radovilsky Z, Eyck JT (2000). "Forecasting with Excel" The Journal of Business Forecasting Methods, Syst .19(3):22-27.

Ragan J, Andrew M, Hadley J, Alexander P. Raymond (2008). "Star Electronics, Inc.: An Excel Based Case Using Financial Statement Analysis To Detect Fraud" J.Bus. Case Studies.4(3):53-70.

Ragsdale C(2007). Spreadsheet Modeling and Decision Analysis. 5th ed. Thomson/South-Western, Mason, OH

Varma TN, Khan DA (2013). Fraud Detection in Supply Chain using Excel Sheet. Inter. J.Computer Applicat.80(2), 20-25.

Appendix I.

Excel Functions	
Version	**No.**
Excel 5.0	299
Excel 2002	40
Excel 2003	0
Excel 2007	5
Excel 2010	61
Excel 2013	50
Total	455

Appendix II.

Functions in Excel 2013	
Category	**No.**
Compatibility	38
Cubes	7
Databases	12
Date and times	24
Engineering	54
Financials	55
Information	20
Logical	9
Lookup and references	19
Math and trigonometry	79
Statistical	101
Texts	30
User defined with add-ins	4
Webs	3