AN ANALYSIS OF SPREADSHEET MODELS USED FOR DECISION SUPPORT

PAUL CRAGG
UNIVERSITY OF WAIKATO, HAMILTON

PETER GRANT
TELECOM CENTRAL, HAMILTON

ABSTRACT

Spreadsheet models are increasingly being used in decision making within organisations. With questions about the quality of these models, an investigation was conducted into the spreadsheet practices in ten firms, with an emphasis on the process of building spreadsheet models. The study showed that spreadsheet models were usually built in an informal, iterative manner, by people from all organisational levels. These people had received very little training in the building of models, which could help explain why 25% of the models contained errors. Other problems were also found. It was evident that the spreadsheet practices in the firms were inadequate. There is a need for increased training as well as setting and enforcing organisational spreadsheet standards. Although the study was viewed as exploratory, it indicated a need for further study into the effect of formal design practices on such factors as the incidence of errors and model creation time.

1. INTRODUCTION

Spreadsheet packages are a very popular part of the rise in end-user computing. Cragg et al (1988) and Cavana (1989) have shown that spreadsheets are used for decision support in various settings by many managers. For many of the managers studied by Mason and Keane (1989) a spreadsheet package was their only decision support tool. However, Ronen et al (1989) point out that many of these individuals lack formal training in systems analysis and design. They surmise that this lack of training leads to systems which are unreliable and incorrect.


Spreadsheet models have great potential for decision support. The problem though, is that an increasing number of decisions could be based on incorrect output from
spreadsheet models. The challenge is to discover how spreadsheet models can be designed and built in order to avoid, or at least minimize, errors while still allowing construction within a reasonable time frame.

This paper discusses an attempt to discover more about the practice of spreadsheet modelling by end users, with particular emphasis on model construction.

2. **METHOD**

Information on how spreadsheets are constructed was obtained from questionnaires, interviews and analysis of existing spreadsheets; while information on how it is felt they should be built was obtained from existing literature.

Ten firms were investigated. These firms varied greatly in size. Within these firms people from all levels (from clerical to top management) were involved with the study. The models varied in size from small (150 cell) models to large, multi-spreadsheet models. The application of these models also varied (from a simple cashflow analysis through to a corporate budget consolidation).

A total of twenty spreadsheet models were investigated. In most cases, copies of the spreadsheet models were obtained for analysis. When copies could not be obtained, interviewee recollections were used to obtain the required information.

**Qualities of Good Spreadsheets**

The existing literature on spreadsheet modelling helped form a framework for questions and analysis. Ronen et al (1989) suggested that good spreadsheet models should exhibit the following qualities:

- they should produce reliable results which are correct and consistent;
- they should be capable of being audited (i.e. others should be able to understand the steps involved in generating various outputs, and thereby understand the model and verify its findings);
- they should be capable of being modified easily without introducing errors; and
- they should be comprehensible (i.e. it should be easy to understand the model and its assumptions as presented in the spreadsheet).

The literature indicated the following types of spreadsheet errors that could occur.

1. **Rounding** - Batson (1986/87) believes that the rounding of output can cause errors, particularly when large figures are involved.

2. **Erroneous formulae** - Brown and Gould (1987), Chussil (1988), and Ronen et al (1989) found errors within formulae to be the most common source of errors in spreadsheet models.

3. **Omitted factors** - Chussil (1988) found that the omission of key factors (from formulae) was a problem with models.
4. Input errors - Mason and Keane (1989) believe that data integrity is usually poorly served by spreadsheets, resulting in erroneous output.

5. Incorrect ranges - the last major error highlighted by current literature was the incorrect specification of ranges. This was discovered by Ronen et al (1989), Williams (1987) and Miller.

In addition, a good spreadsheet model should be suited to its users, meet with its specifications, and produce useful and meaningful output.

3. MODEL USE AND QUALITY

Length of Model Use

While some of the models investigated had only been in use for a month, 10 had been in use for more than 6 months, and 6 for a more than a year. One model had been used since 1983.

It appears that the application remains fairly unchanged over time. Therefore, models do not have to be completely rebuilt, but rather modified from time to time.

Spreadsheet Modeller

Six of the respondents were in middle management positions. Only one was in a top management position. The remaining respondents were spread evenly among clerical, supervisory, and lower management positions.

Typically, training had been conducted in house. However, these courses were generally short and lasted for only one or two days. Only one third of the model builders had had any training in programming.

Spreadsheet Standards

Only one of the organisations investigated had formal spreadsheet standards. These standards were presented in a booklet and given to those involved with spreadsheet models.

Number of Modifications

While one third of the spreadsheets investigated were in their first version, it was found that the majority of spreadsheet models were updated continuously. In fact, one model was up to its 60th version.

On average the models had been updated 6.5 times. The reasons for modification indicated that spreadsheet models tend to be built without much thought to specification. This lack of specification had led to the need for subsequent modification.

Layout

Although the majority of the spreadsheet models analysed were found to have some separation of sections (eg input and output), only 55% had a clear separation (the others had sections of input and output scattered throughout the model).
No model separated the calculations from the rest of the model. The calculations tended to be mixed with the output.

The lack of proper format in the remainder of the models made them very difficult to understand. This would make future modification by others (and even the modeller) very difficult and dangerous.

Errors Found

Of the models analysed, 25% were found to contain errors; all involved formulae. However, because some models were very difficult to understand, it is possible that a more thorough analysis could have discovered more errors.

Assumptions

In general, assumptions were well labelled and separated from the remainder of the model. In fact, 60% of the models presented their assumptions in this way. This is very encouraging as it reduces errors during modifications.

Data Validation

Input data had been checked for 80% of the models. This was either manually (prior to entry) or by the model itself. Most relied on the user checking the figures manually. Four of the models used range checking to test data validity. One model used its report macro to check selected values and produce an error report.

Cell Protection

Cell protection allows cells to be 'locked' so the user cannot overwrite their contents. This facility was only used by 30% of the models investigated.

4. THE BUILDING PROCESS

All of the models investigated were built using relatively informal methods, i.e., without formal specification, design, coding, or documentation. The majority were entered straight onto the computer without any prior preparation. Most of the respondents commented that they knew what they wanted and therefore simply typed it into the computer.

The one factor common to most of the spreadsheets was an iterative stage at the end of the process. This iteration involved backtracking to modify the model for various reasons (such as errors in formulae, cosmetic changes, and additions to the model). On average, the models had been modified 7 times.

Only 5 of the modellers defined the model's input and output prior to building. Two modellers defined the calculation rules prior to coding.

The following is a summary of findings with regards to the various stages in the formal design process

Specification

Only 7 of the spreadsheet models involved specification by others. Although this is low, it can be attributed to the fact that most of the models were used solely by their creator. Two had been built for others without any input from them.
Design

Design took place with only 8 of the models. Even this design was very rudimentary, mainly involving the design of input screens and reports. Planning the layout of the model only took place with 8 of the models studied.

Although some of the models were fairly simple, the lack of design is probably responsible for the 84% modification rate and the fact that 6 of the modellers admitted that there had been problems with their models. In addition, the fact that most of the respondents said that they had to modify existing cells in an iterative manner, could probably be attributed to poor design.

Testing

All but one of the spreadsheet models were tested. The majority of the models were tested by putting data into the model and working through the formulae manually to check the output.

Historic data was used to check 7 of the models while cross checks were used by 5 of the models. Only one model was tested using a spreadsheet audit package (a package which helps alert the modeller to errors by listing such things as references to empty cells.).

Almost all of the models were tested in one way or another. However, model validity would be increased if the models were tested by someone other than their author. This was done in only one case.

Documentation

Documentation of the models was poor. Only half of the models contained documentation of any kind. Of those that did have documentation, the majority only provided sample input and output which was obtained from a printout of the entire model.

5. CONCLUSIONS

This study of twenty spreadsheet models lends real support to the concerns expressed by others about the validity of many spreadsheet models. A quarter were found to contain errors. In addition about one half had poor layout, and most relied on the users to check the validity of data input. Only one third used cell-protection. Over half of the models were built with no prior design or planning. Instead, they were just typed into the computer. On the bright side, all but one model had been tested, though only one by another user, and another using a spreadsheet audit package.

It is evident that current spreadsheet practice is inadequate. Current literature provides many methods for the reduction of spreadsheet errors. Most of these are easy to implement. In general, there is a need for increased training and the setting of organisational spreadsheet standards.

REFERENCES


Mason, D. & Keane, D. "Spreadsheets: Solution or Problem", Interface, October 1989, pp.82-84.

Miller, S.E. "8 Ways to Avoid Worksheet Errors", reference unknown.

