2006

Enhance your own research productivity using spreadsheets (From sin to salvation)

Janet Wagner
Jeffrey Keisler, *University of Massachusetts Boston*

TUTORIALS IN
OPERATIONS RESEARCH
INFORMS—Pittsburgh 2006

 | ISBN 0000-0000
DOI 10.1287/educ.1053.0000

Chapter 1

Enhance Your Own Research Productivity Using Spreadsheets

Janet M. Wagner
Department of Management Science and Information Systems, University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA, 02125, USA janet.wagner@umb.edu

Jeffrey Keisler
Department of Management Science and Information Systems, University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA, 02125, USA jeff.keisler@umb.edu

Abstract Spreadsheets are the modeling tool of choice for a large number of OR/MS researchers. Surveys of users show that most do not use basic good practices and most large spreadsheets have flaws leading to results ranging from wasted time to downright scandal. Fortunately, many of the solutions to these problems are already known and easily learned. This workshop, taught by OR/MS modelers who have first hand experience with both "sin" and "salvation" in the spreadsheet kingdom, presents the authors' "top ten" Excel methods and four major spreadsheet applications from their own research and professional lives. Tutorial participants, bring your laptops!

Keywords tutorials in operations research; spreadsheets

1. Introduction

Like Rodney Dangerfield, spreadsheets don't get no respect. Casimir [5] proclaimed "Real Programmers Don't Use Spreadsheets". Grossman et. al. [10] describe multiple examples showing a "perception that spreadsheets are somehow different than other programming tools, and that spreadsheets are suitable for personal use but not for important tasks which are reserved to information systems".

However, the use of spreadsheets is ubiquitous in both business and OR/MS. Microsoft Excel alone has an installed user base of 440 Million licenses [14], with additional hundreds of millions using Open Office, Quattro Pro, Lotus 123, and Gnumeric. Scaffidi et. al. [23] estimates that the number of spreadsheet and database users in the U.S. alone will reach 55 million in 2012, over four times their estimate of 13 million "professional" programmers. Evidence is growing about the many uses of spreadsheets for critical business processes. For example, the paper "Stop That Subversive Spreadsheet" by Butler and Chadwick [4] describes the nexus of concerns of both academicians and practitioners that led to the formation of the European Spreadsheet Risk Interest Group (EuSPRIG) [9]. As just one example, Croll [6] talks about the ubiquitousness of spreadsheets in the London financial community (called the "City of London"), and concludes "it is completely within the realm of possibility that a single, large, complex but erroneous spreadsheet could directly cause the accidental loss of a corporation or institution, significantly damaging the City of London's reputation." Estimates of the number of OR/MS spreadsheet users are harder to come by. However the extent of the coverage of spreadsheets in OR/MS textbooks and the existence of groups such as EuSPRIG and, within INFORMS, of the Spreadsheet Productivity Research Interest Group (SPRIG) [26] provide evidence that spreadsheets are a common tool for those in OR/MS fields.

The focus of this tutorial is specifically on the use of spreadsheets as OR/MS application development tools. The goal of this tutorial is not just to develop spreadsheet examples similar to those available in a comprehensive Excel manual but rather to gain an understanding at an abstract level of what spreadsheet tools are and how to relate them specific OR/MS modeling needs. In this tutorial, we will provide concepts and methods for building, verifying, and using spreadsheets in a way that maximally enhances productivity. We will also present examples of spreadsheets, developed and used in the authors' professional lives, to both model good spreadsheet practice and to illustrate our concept of matching spreadsheet tools to real professional OR/MS needs.

2. Spreadsheets: From "Sin" to "Salvation"

Spreadsheets can be almost too easy to use. It is quite possible for OR/MS models to push spreadsheets to (and beyond?) the limits of their capabilities. Have you ever built a large, complex spreadsheet model, which ended up taking you more time to debug than the original development time? When you revise an article after six months, do you have to spend large amounts of time remembering exactly how your spreadsheet works? Is there a significant chance your model is actually invalid?

EuSPRIG [9] maintains on its website press accounts of important spreadsheet mistakes; there were 85 such stories at the time this tutorial was written. Recent examples include the City Council of Las Vegas having to postpone their vote on the city budget because of over five million dollars of errors in the spreadsheet output provided as part of the budget bill, and several examples of companies having to restate earnings by millions of dollars due to "clerical errors" in spreadsheets. Striking in this archive is the magnitude of the effects of the reported mistakes and the fact that despite the magnitude and criticality of these applications the mistakes occur mainly from simple common mistakes such as botched sorting or misspecified sum ranges. We would all like to keep ourselves and our spreadsheet exploits out of the EuSPRIG error archive (and the press), but undoubtedly so did the authors and users of those reported incidents.

The challenge, then, is that we are all "sinners" regarding robust and rigorous spreadsheet design and implementation. In this tutorial we will explore the path of "salvation" with specific attention paid to certain paving stones along that path. We believe that like any other information system application, spreadsheets pose risks. However, there are many straightforward techniques that can help reduce and manage those risks. The opportunities of spreadsheets are simply too great to dismiss this technology completely, even when developing complex systems.

3. Sources of Salvation (Background knowledge)

Strategies for the effective and efficient use of spreadsheets can be drawn from a number of areas, including software development and engineering, OR/MS modeling, the psychology of error, and traditional auditing. In addition, commercial applications to assist with spreadsheet development and use appear on the market almost daily. We will give some selected representative sources for these background areas below. We also recommend both the EuSPRIG [9] and SPRIG [26] websites, which maintain links to a variety of research articles, conference presentations, books, and products related to spreadsheet modeling and development.

Software development and engineering: current spreadsheet practice has been compared to the "wild west" days of early programmers. The disciplines and methods of the field of software engineering, which have helped to tame the development of conventional software, have much to offer spreadsheet developers as well. Boehm and Basili [3] provide data that show "disciplined personal practice can reduce deficit introduction rates [in programs] up to

75 %." Textbooks and reference works on software engineering include those by Sommerville [25], Pressman [18], and McConnell [13] [12].

OR/MS modeling: spreadsheet applications of OR/MS models and techniques have become an integral part of many textbooks and reference books. Multiple examples can probably be best obtained in the exhibit halls accompanying this conference, but "classics" would include books by Ragsdale [21], Winston and Albright [1], Hannah, Ahuja, and Winston [24], and Powell and Baker [17]. Tennent and Friend [27] is another useful book, written for economists.

Psychology of error: humans make errors, and psychologists among others have studied factors that can lead to either more or less of them. Ray Panko maintains a webpage [20] with a comprehensive bibliography on both human error in general and spreadsheet errors in particular.

Traditional auditing: the process of reviewing the accuracy of financial statements has much in common with processes for reviewing the accuracy of spreadsheets. Basic textbooks on auditing include those by Rittenberg and Schwieger [22] and Arens *et. al.* [2]. The previously mentioned SPRIG website [26] contains a listing of available packages for spreadsheet auditing. O'Beirne [16] is a useful spreadsheet-oriented book, covering auditing as well as spreadsheet design topics.

4. Process and Principles for Salvation (Spreadsheet Design and Engineering)

Paradoxically, research productivity using spreadsheets is probably most enhanced by *investing* time — as long as that time is spent before touching a keyboard. Following Powell and Baker [17] we advocate following a thoughtful process for spreadsheet development, with separate phases of spreadsheet design, building, and testing. As Powell and Baker point out, builders don't build buildings without blueprints and neither should researchers build spreadsheets without plans.

Principles adapted from Powell and Baker for ease-of-use and for avoiding the dreaded "spaghetti code" include:

- Separating data from calculations, and separating analysis from presentation
- Organizing spreadsheets with a logical progression of calculations (top to bottom, left to right)
- Developing data and analytical "modules" (including grouping within worksheet, and the worksheet structure itself)
- Sketching, in advance of development, major spreadsheet elements and calculation flow
- Using graphical aids to modeling (we are particular fans of influence diagrams)
- Giving thought to and consulting with the end users of the spreadsheet on their needs (the user, who is not necessarily the spreadsheet builder, may have a very different view of the process than the spreadsheet analyst)
- Keeping formulas short and simple
- Planning for documentation "as you go"
- Stating model assumptions explicitly
- Using formatting aids, such as color, text differences, cell outlining
- Protecting end-users from unnecessary analytical details, and inadvertent changes

In Excel, basic built-in tools supporting these principles include:

- Availability of absolute vs. relative references
- Cell and text formatting
- Protected and locked worksheets and cells
- Data (range) names
- Function wizards

We assume readers are familiar with these basic tools, although we will quickly go over them as requested in a "hands-on" manner in the tutorial session. Readers unfamiliar with these Excel elements can explore their use using the built-in help, a basic Excel text [11], or in Baker and Powell [17]. (Or, of course, using the time-honored approach of asking a friend.)

We would also suggest that an investment of time exploring these basic tools, in advance of any research or modeling efforts, is likely to pay multiple dividends. Both of the authors have spent time systematically examining all the available functions and all the cell and text formatting options in Excel. We found this investment of time exploring spreadsheet capabilities is repaid many times over from the new ideas and possibilities for their application that we gain from it. Walkenbach [28] is a comprehensive Excel book, favored by the authors.

5. On the Path to Salvation (Advanced Tools)

More advanced (and lesser known) Excel tools are available that, if properly and consistently used, can aid in the efficient and effective development and use of research and end-user spreadsheets. In this section we will give some "step-by-step" directions as well as hints on the use of the following Excel methods:

- Comment and formula display options
- Data validation
- Spreadsheet auditing
- Built in error checking

Note: material in bold describes MS Excel (Office 2003) commands.

Comment and formula display options: A text comment to accompany a cell is added by: **Insert - Comment**. Comments don't have to clutter up the spreadsheet, since the default is to show them only when the cursor is on the particular cell. (A cell with comments is indicated by a red triangle in the corner of the commented cell.) Comments are a good way to document calculations so a given formula is understandable six months from now. **Tools - Option - View** gives different options. A comment can be removed by: **Edit - Clear - Comments**

To see a formula and a color coded display of the cells referenced in the formula double click on the cell, or use F2. All formulas in a worksheet can be displayed simultaneously by pressing **Ctrl** + ˜ (tilde).

Data validation: If users enter data into a spreadsheet, guidance can be provided to them (and errors avoided) by using: **Data - Validation**. When data validation is required for a cell, the value can be restricted (e.g. "between 0 and 50") as can the type of value (e.g. "whole number"). Data validation menu items also allow comments to be specified that will show when the cell is selected as well as the error message that will appear when the data is not entered according to the specifications. **Data - Validation - Clear All** removes the validation specifications.

Spreadsheet auditing: Excel comes with built in formula auditing functions, which are accessed by: **Tools - Formula Auditing - Show Audit Toolbar**. These auditing functions are particularly helpful in parsing and testing complex formulas. The audit toolbar has tools that graphically trace the cells used in a formula (**Trace Reference**s), or trace where a particular cell is used in a subsequent formula (**Trace Dependents**). Another useful function in the audit toolbar is **Evaluate Formula** which shows steps in a complex formula calculated a piece at a time.

Error Checking: starting in Excel 2002, Excel looks for certain basic errors in formulas. We note that, like spell and grammar check in word processing programs, some people find these checks more annoying than helpful. **Tools - Options - Error Checking** brings up

a menu which allows adjustment for what errors are and are not looked for (and/or turn error checking completely on or off, as wished).

All of the above tools are general purpose tools that will enhance the development process for all spreadsheets. We maintain that due to the complexity of most OR/MS models, building on a solid platform of good spreadsheet practices is particularly important. Models with frequent comments for complex formulas, which have had their formulas audited, have been error checked, and with built in data validation will most likely be able to be followed six months from now, can turned over to a successor with ease, and we will be easier to test and use.

6. The End of the Road: Putting it All Together (Techniques and Applications)

The focus of this tutorial is to find a mathematically elegant way to use the structure and functionality available in spreadsheets to encode the structure of your problem. In this section we'll go over our "top 10" set of Excel methods for OR/MS researchers. We will motivate this list by showing examples of how we have combined these specific "top 10" tools, and the more general good spreadsheet design principles discussed in previous sections, into "killer aps".

We start by observing that Excel methods can be classified broadly as "interface" tools and "analysis tools". Most applications will require both types of tools, but the balance of these two functions will vary by the application and intended use. A spreadsheet intended mostly to answer a research question may focus mainly on the analytical tools with little attention to interface/presentation, while another system intended to support non-technical decision makers may require mainly interface tools. Thought, however, needs to be given to both functions — no matter the application.

6.1. Interface Tools

6.1.1. How we Doin'?: *A spreadsheet for nationally-normed student survey results* This application came from one author's foray into college administration, where an OR/MS sensibility infused (for good or for ill) the position of Associate Dean. The value of this spreadsheet application is in its ability to present large amounts of data in a compact and engaging form. The file is available as studentsurvey.xls. Note the data in this spreadsheet has been altered, both for UMass Boston and the benchmarking information. The values in this example are representative values, not the actual ones.

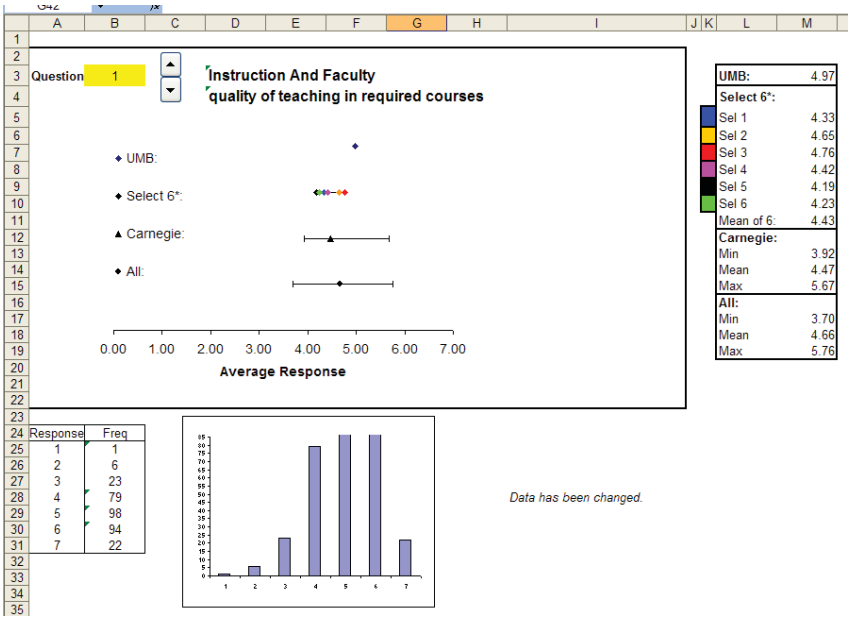
The College of Management at UMass Boston, like many AACSB accredited schools participates in student assessment surveys using an instrument from Educations Benchmarking Inc. [8]. EBI surveys have the advantage not only of providing the responses of our own students, but providing national benchmarks as well (and comparison data for 6 other benchmark institutions). EBI provides multiple analyses and presentations of results but we found it difficult to both interpret and distribute the results of these surveys. The spreadsheet presented here, provides an interactive graphical representation for each of the 66 survey questions, showing in one compact user-friendly display UMass Boston's results, compared to the six benchmark schools, the set of schools in our same Carnegie classification, and the entire set of schools using EBI that particular year (see Figure 1).

This first example relied heavily on the interface focused tools of:

(1) Conditional Formatting
(2) Graphs
(3) Form Controls

Method 1: Conditional Formatting Excel allows the user to change the formatting of what is displayed in a cell, depending on the value (or formula) in the cell, a feature accessed by:

FIGURE 1. Student Survey Display



Format - Conditional Formatting. The resulting menu allows the user to set one (or more) conditions that will then lead to a specified format (including cell background, font, and cell outlining).

Method 2: Graphs The ability to simultaneously present information in text, numbers, and graphs is one of the major sources of the power of spreadsheets. The graphical features of Excel can be accessed by: **Insert - Chart** (or by clicking on the chart icon in the standard toolbar). This tutorial does not include a comprehensive discussion of all the Excel graph options. However, we want to highlight that particularly interesting interfaces can be created by using "live" graphs, which respond to changes in user input data.

It should be noted, there are also significant limitations to Excel graphs, particularly for more scientific and other technical uses. Multiple graphing computer packages exist, which should certainly be considered for specialized uses.

Method 3: Form Controls A set of interesting Excel controls can be revealed by: **View - Toolbar - Forms**. We will focus on the two tools of **Scroll Bar** and the **Spinner**. Both of these controls are accessed by clicking on the appropriate icon, dragging an appropriately sized area on the spreadsheet itself, right clicking on the control, and then choosing **Format Control**. These tools allow a "kinesthetic" way to enter or change data, and can be particularly useful in creating applications designed to engage the spreadsheet user in data manipulation. (We're not sure why moving a bar or clicking a little triangle is so much more engaging that retyping a number, but from experience we know that these tools really do draw users in.)

We would encourage the reader to open and explore the first spreadsheet (studentsurvey.xls) here. Tools that were used to produce this spreadsheet include lookup functions (discussed below) and the creative use of formatting, form controls, and graphical functions of Excel. The graph is based on an x-y plot, with three separate data series. Note that some data are hidden (white text, in columns N through Q). The horizontal lines on the plots come from the error bar capability. The spinner is used to pick a question, which looks up the data for that question (both from the internal survey results and the national norms), and the graph then automatically re-displays.

This spreadsheet was given to all the standing committees in the College, which included most of the tenure track faculty. Discussions that semester, involving multiple staff and

faculty groups, provided ample evidence that this spreadsheet was used by multiple people. The information gained from this effort resulted in several changes to existing procedures and new initiatives. At least partly as a result of these programmatic changes, when another survey was taken two years later, the undergraduate results improved on 63 out of the 66 questions.

6.2. Analytical Tools

6.2.1. Classrooms need chalk and students: *What class schedules can tell you*

The second example is also a simplified version of a "real" spreadsheet, again used for college administration. The file is available as: classsched.xls. Again, this spreadsheet contains representative data, not any actual semester's schedule. This spreadsheet involves some important but fairly simple calculations, however its real value is again its ability to present data in usable form. It started as a single-purpose spreadsheet, to calculate faculty deployment ratios (e.g. percent of MBA credits presented by full-time faculty) required by AACSB — using basic information supplied by the registrar's class schedule and the college's faculty list. However, once this data set existed, questions were posed about these data that had never been imagined. Hence, this spreadsheet is one that developed over several years, with a new report being created each time someone else said "could you tell me . . . ?". In this case, the presentation issue is that data available from the run-of-the-mill class schedule has multiple uses and needs to be displayed in multiple ways.

The second example, is based on the analytically focused tools of:

(4) Lookup Functions
(5) Sorting
(6) Filtering
(7) Pivot Table

Method 4: Lookup Functions Lookup Functions: The lookup and reference functions are some of the most useful Excel functions in creating high functioning spreadsheet systems. We will focus on the HLOOKUP and VLOOKUP functions, but all of the lookup and reference functions are worth a look. These functions can be accessed by: **Insert - Function** (or from the **fx** icon). The HLOOKUP function is used to look up a value across a row; the VLOOKUP function is used when you are looking for a value down a column. Among other uses, these functions can be used to obtain functionality similar to a relational database. They can also enable data to be entered in a logical and compact form, so that entries can be built up from components instead of having to retype data multiple times. For example, to compile a list of faculty members, one can use a LOOKUP function to determine what college a given department is in instead of having to remember and type it each time.

Method 5: Sorting Before we discuss this method, we need to point out that sorting is a double-edged sword. The ability to sort information, by rows or by columns, is both one of most useful (and used) Excel capabilities *and* is also a way to cause really serious errors. Sorting capabilities are accessed by selecting the range containing the data to be sorted then **Data - Sort**. Where errors commonly occur is in selecting the incorrect range of data to be sorted. Sorting should be done with care. If one was to sort all but one column of a given table the error can only be corrected using the "undo" function, which means if the error is not caught quickly it may not be fixable at all. Using named ranges for data that are to be frequently sorted is a good way to reduce the occurrence of such errors.

Method 6: Filtering and Subtotals Filtering allows the user to choose a subset of a data range, according to a user-defined criteria, for data organized in columns with column headings. Filtering is accessed by: selecting a column label (or labels) and then **Data - Filter - AutoFilter**. Small triangles then appear at the top of the columns. Selecting the triangle shows a list of values in the column; clicking on a value filters for that value. More advanced

custom filters can be created with other menu options. The triangles can be turned off (and the full unfiltered set of data restored) by repeating **Data - Filter - AutoFilter**.

Helpful to use with filtering is the SUBTOTAL function, which we find useful if rather non-intuitive. Subtotal has two arguments, the first being a number which defines the calculation (use 9 to get a sum) and the second being the data range to be used in the calculation. When no filter is applied, SUBTOTAL works like whatever function the user chooses (so with 9, Excel would calculate a regular sum). However, when the data is filtered, SUBTOTAL only calculates the chosen function for the displayed value (e.g. shows a subtotal).

Method 7: Pivot Table In a way, Pivot Tables are an extension of the subtotal function. For example, suppose a user had a list of employees, with associated departments and salaries. Once could manually construct a table of total salary budget by department by using the filter and the SUBTOTAL function to choose each department in turn and then recording that department's total salary. The pivot table function, however, will create this table automatically.

A pivot table works only on data arranged in columns with a column label entered for every column. The Pivot Table is accessed by: **Data - PivotTable and PivotChart Report**. The first two menus are fairly self-explanatory; at the third click on **Layout**. Here one has a chance to set up a table. Variables that the data are to broken down by are dragged to the row or column area. (So in the departmental salary example, the department would be put in the column space.) The values to be broken down (salaries in the example) are dragged into the data area, and by clicking on the label in the data area the calculations to be performed can be changed. To filter what values get into the Pivot table, other variables can be put into the page area. Click **OK** then **finish** and the breakdown (or pivot) table will appear.

Pivot tables are a very rich resource, and there is more to them than can be explained in this short tutorial. Chapter 21 of Walkenbach [28] discusses pivot tables in more detail. We have found that pivot tables are another example of a function that once a user grasps the basic idea, much of the rest can be picked up by playing around with them.

We would encourage the reader to open and explore the second spreadsheet (studentsurvey.xls) here. The spreadsheet for this second example was designed using the "good spreadsheet practice" of providing compact, logically organized data, followed by (separate) analyses, followed by (separate) presentations of the results. After the first worksheet, which provides a "front page" to the spreadsheet (see Figure 2), the next three worksheets are data (course list, instructor list, and then class sections). Filtering, sorting (macros attached to buttons using simple VBA code), and lookup functions help keep the data compact and organized, and reduce errors by drastically reducing retyping (and allowing quick, and reliable data changes). The next worksheet (see Figure 3) includes the pivot tables that are necessary for the ratio analysis. Since these pivot tables are used only by the analyst, no particular attempt was made to make them user-friendly. The following sheets focus more on presentation, covering a wide range of uses and presentations. As well as a managerial presentation of the ratio results, reports exist to show scheduling (which nights are MBA classes offered, see Figure 4), faculty workload (number of courses and total students, see Figure 5), a more user friendly presentation of the class schedule, and a report to ensure that nobody is double-scheduled (which, from sad experience, turned out to be important to check).

This system for semester class scheduling has been used for more than five years. It is used prospectively (as the semester schedule is being determined) and retrospectively (to provide historical reports). The spreadsheets are available on the internal college servers, and are used by the college's administration (Associate Dean and MBA Director), as well as by the Department Chairs, and the clerical staff. It is part of how the college does business. We believe that the widespread use of this system has occurred because each user can access (and manipulate) these data in exactly the way they like and need to interact with them.

FIGURE 2. Class Schedule Front Page

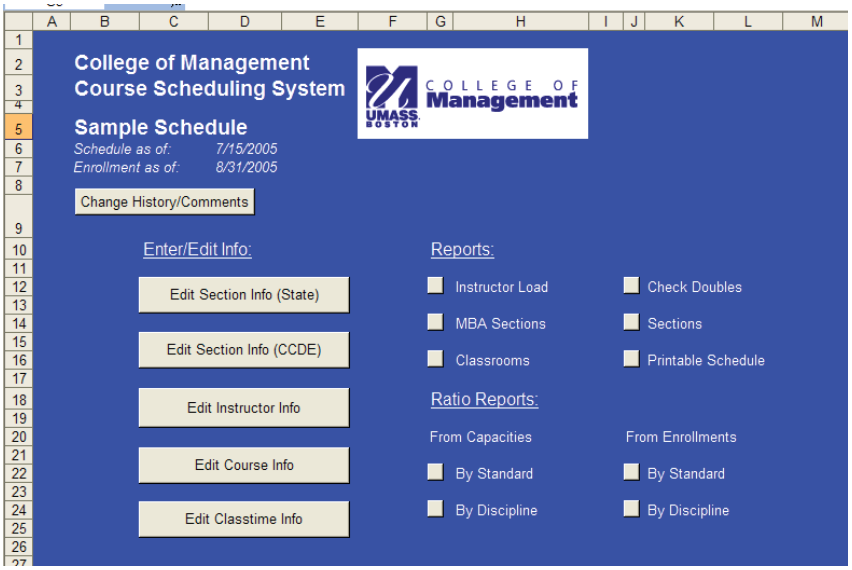
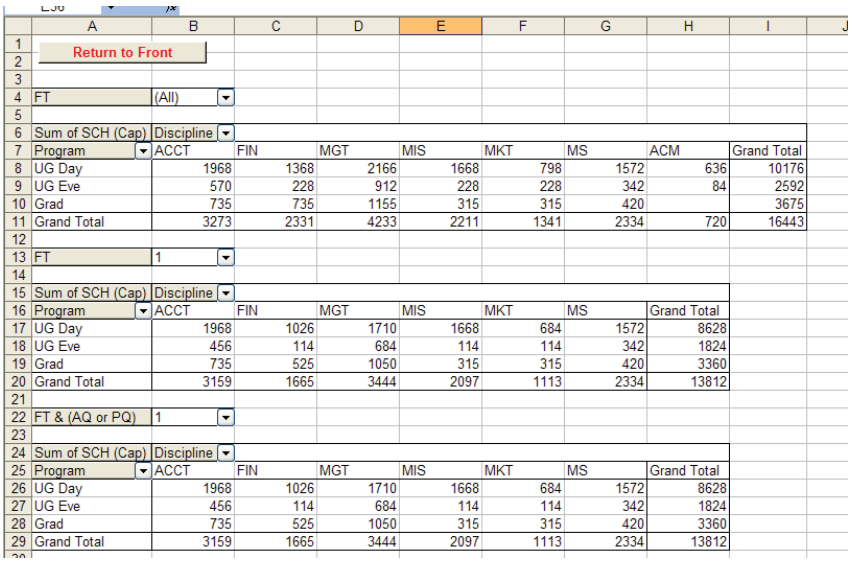


FIGURE 3. Class Schedule Pivot Tables



Return to Front

FT	(All)

Sum of SCH (Cap)	Discipline							
Program	ACCT	FIN	MGT	MIS	MKT	MS	ACM	Grand Total
UG Day	1968	1368	2166	1668	798	1572	636	10176
UG Eve	570	228	912	228	228	342	84	2592
Grad	735	735	1155	315	315	420		3675
Grand Total	3273	2331	4233	2211	1341	2334	720	16443

FT	1

Sum of SCH (Cap)	Discipline						
Program	ACCT	FIN	MGT	MIS	MKT	MS	Grand Total
UG Day	1968	1026	1710	1668	684	1572	8628
UG Eve	456	114	684	114	114	342	1824
Grad	735	525	1050	315	315	420	3360
Grand Total	3159	1665	3444	2097	1113	2334	13812

FT & (AQ or PQ)	1

Sum of SCH (Cap)	Discipline						
Program	ACCT	FIN	MGT	MIS	MKT	MS	Grand Total
UG Day	1968	1026	1710	1668	684	1572	8628
UG Eve	456	114	684	114	114	342	1824
Grad	735	525	1050	315	315	420	3360
Grand Total	3159	1665	3444	2097	1113	2334	13812

6.2.2. Up and About: ***Calculation of seasonal indices on top of a general linear trend*** The third example may be most useful as a teaching example (one of the authors remembers being shown a version of this example at a Teaching Management Science workshop). It is also a good example of the functionality that occurs from the creative exploitation of the flexibility in spreadsheets. The file is available as: seasonal.xls

A common forecasting method involves developing time-series model with a linear trend and seasonal indices. The example in the spreadsheet involves U.S. Commerce Data (Survey of Current Business) on quarterly general merchandise sales (in millions of dollars) from 1979 to 1989 (obtained from DASL [7]). An example such as this would traditionally be used in a class on business statistics or operations management.

This example relies on the analytical focused tools (probably familiar to most OR/MS professionals) of:

FIGURE 4. Class Schedule MBA Schedule Display

	A	B	C	D	E	F	G
1							
2	Return to Front						
3	CourseType	Grad Core					
4							
5	Count of Section	Classtime					
6	Section	M 6:00	TU 6:00	W 6:00	TH 6:00	Sa 9:30	Grand Total
7	AF 601		1				1
8	AF 610	1					1
9	AF 620				1		1
10	MGT 650			1	1	1	3
11	MGT 660		1				1
12	MGT 689		1				1
13	MKT 670	1					1
14	MSIS 600	1					1
15	MSIS 630		1				1
16	MSIS 635			1			1
17	MSIS 640	1					1
18	Grand Total	4	4	2	2	1	13
19							
20							
21							

FIGURE 5. Class Schedule Faculty Workload

	A	B	C	D	E	F
1						
2	Return to Front					
3						
4						
5						
6	Dept	AF		Dept	MM	
7						
8	Sum of SecEq			Sum of SecEq		
9	Name	Total		Name	Total	
10	Brian	3		Alton	1	
11	Charles	3		Coppers	2	
12	Chap	3		Dorn	2	
13	Core	3		Eve	2	
14	Fred	2		George	2	
15	Graham	2		Gund	3	
16	Holt	3		Hardy	2	
17	Hughes	3		Henry	1	
18	Jackson	1		Lars	2	
19	Johnson	2		Lind	3	
20	Jones	3		Lorne	1	
21	Overton	3		McGowen	1	
22	Ryan	3		Mercy	3	
23	Varga	3		Naby	2	
24	Vegas	4		Powell	1	
25	Worthen	2		Rorn	3	
26	Zabby	4		Rundi	4	
27	Grand Total	47		Samuels	2	
28				Wall	2	
29				Werth	4	
30				Wurnd	3	
31				Zaylor	4	
32				Grand Total	50	
33						

(8) Statistical add-ins (e.g. regression)
(9) Solver

Method 8: Statistical add-ins Excel has a number of built in statistical functions which can be accessed by: **Tools - Data Analysis**. (Note, the data analysis pack is not always part of the standard installation procedure for Excel, and may have to be added in later.) Multiple statistical functions are available, most of which have fairly easily followed menus. Note

that Excel is not a special purpose statistical package, and so is not considered as robust as several commercially available statistical packages. Some of the more advanced functions have, at least in the past, had errors for example with the handling of missing data. (See [15] for a report on Microsoft's responses to these issues.) Nonetheless, as part of a larger system, the ability to include statistical analysis with other types of calculations can lead to Excel as the statistical package of choice.

Method 9: Solver Again, it is beyond the scope of this short tutorial to go through all aspects of Solver. Solver is also an Excel add-in, and can be accessed by: **Tools - Solver**. The user must specify the cell containing the objective value (the Target Cell), the decision variables (the Changing Cells), and the constraints (added one-by-one). The option screen allows the user to choose the solution method (linear, types of non-linear, etc.). Solver is thoroughly discussed in several OR/MS textbooks such as [21] [1] and [24].

The first worksheet (see Figure 6) calculates seasonal indices using the two-step Seasonal Index Method (*c.f.* [19], Chapter 12). First a linear regression is run on the original data and used to calculate a predicted value for each quarter. Then the ratio of the actual data to the predicted amount is calculated, and these ratios are averaged for each individual quarter. These average ratios are then used as the seasonal indices, and the seasonalized prediction is then calculated as the predicted linear regression value multiplied by the seasonal index. The first worksheet uses the statistical add-in for regression.

FIGURE 6. One-Step Linear/Seasonal Calculations

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Original Data				Traditional Seasonal Analysis											
2	TIME	Quarter	GMER		Linear Pred	Actual/Fore	Seasonal Forecast	Linear Only Error-Sq	Seasonal Correction Error-Sq							
3	1	1	30363		27053.6	1.122	21606.95	10952027.61	76668360.29			**Regression**				
4	2	2	25812		27534.2	0.937	26225.29	2965837.396	170804.5956			SUMMARY OUTPUT				
5	3	3	26026		28014.7	0.929	26055.47	3954952.076	868.5139376							
6	4	4	38081		28495.3	1.336	38797.07	91886572.66	512749.9539			*Regression Statistics*				
7	5	1	21909		28975.8	0.756	23142.14	49939620.41	1520646.274			Multiple R	0.599927			
8	6	2	26615		29456.3	0.904	28056.09	8073227.172	2076749.881			R Square	0.359912			
9	7	3	27039		29936.9	0.903	27843.22	8397754.566	646774.2059			Adjusted R S	0.3443			
10	8	4	40170		30417.4	1.321	41414.17	95112555.21	1547956.021			Standard Err	8144.391			
11	9	1	23641		30898.0	0.765	24677.34	52663742.15	1073993.989			Observations	43			
12	10	2	30477		31378.5	0.971	29886.9	812746.0867	348216.2859							
13	11	3	30065		31859.1	0.944	29630.97	3218686.33	188377.7809			ANOVA				
14	12	4	43805		32339.6	1.355	44031.27	131455048.1	51199.02126				*df*	*SS*	*MS*	*F*
15	13	1	24256		32820.2	0.739	26212.53	73344848.09	3828004.765			Regression	1	1.53E+09	1.529E+09	23.053711
16	14	2	30976		33300.7	0.930	31717.71	5404258.596	550133.1725			Residual	41	2.72E+09	66331111	
17	15	3	30579		33781.3	0.905	31418.73	10254415.22	705141.5706			Total	42	4.25E+09		
18	16	4	45471		34261.8	1.327	46648.38	1256462.31	1386212.379							
19	17	1	25891		34742.3	0.745	27747.72	78346263.96	3447411.739				*Coefficient*	*standard Err*	*t Stat*	*P-value*
20	18	2	33050		35222.9	0.938	33548.52	4721442.04	248519.9924			Intercept	26573.07	2527.984	10.511564	3.325E-13
21	19	3	33385		35703.4	0.935	33206.48	5375133.448	31869.65657			TIME	480.5455	100.0839	4.801428	2.12E-05
22	20	4	50311		36184.0	1.390	49265.48	1996727.26.3	1093115.401							
23	21	1	29177		36664.5	0.796	29282.91	56063020.34	11217.48652			**Seasonal Averages**				
24	22	2	36855		37145.1	0.992	35379.33	84140.46998	2177613.864			(Actual/Forecast)				
25	23	3	35418		37625.6	0.941	34994.23	4873564.968	179679.8026			Quarter				
26	24	4	53126		38106.2	1.394	51882.58	225595573.3	1546089.3			1	0.799			
27	25	1	30325		38586.7	0.786	30818.1	68255788.2	243152.1243			2	0.962			
28	26	2	37977		39067.3	0.972	37210.13	1188648.52	588083.2624			3	0.930			
29	27	3	37028		39547.8	0.936	36781.98	6349377.124	60524.06834			4	1.362			
30	28	4	54369		40028.3	1.358	54499.68	205654457.7	17078.51883							
31	29	1	31594		40508.9	0.780	32353.3	79475227.15	576531.2032			**Mean Sqauare Error:**				
32	30	2	37617		40989.4	0.918	39040.94	11373307.06	2027611.666			Linear Only		8547.448		
33	31	3	36297		41470.0	0.875	38569.74	26759710.27	5165327.903			Seasonal Correction		2047.504		
34	32	4	52457		41950.5	1.250	57116.79	110386031.4	21713624.27							

However, the interesting observation is that since regression is, in fact, an optimization method (minimizing the total least squares error) this two-step procedure (regression then smoothing) can be done in one step, resulting in a lower total error than doing the two steps separately. In the example, worksheet 2 (see Figure 7) redoes the seasonal index calculations, using the non-linear optimization capabilities of Solver to find *simultaneously* the coefficients of the linear model *and* the seasonal indices (with the constraint that the seasonal indices add up to the number of seasonal periods, in this case 4). In this case the reduction in total error is not high, but it is nonetheless reduced.

The value in this example is developing, in students as well as in researchers, the creativity (supported by the flexibility of spreadsheets) to view and manipulate problems using a variety of methods. Traditional regression analysis and optimization are not commonly combined in this way.

FIGURE 7. Two-Step Linear/Seasonal Calculations

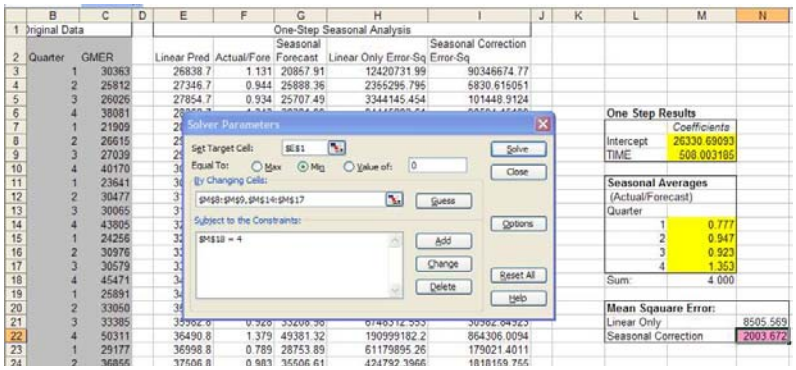


Figure 7: Two-Step Linear/Seasonal Calculations

6.2.3. Make up your mind: *Live Decision Analysis* This fourth example comes from trying to overcome the fact that certain spreadsheet methods such as sorting, pivot tables, and some of the statistical add-ins (regression) are not "live", in the sense that if one changes the data one has to take additional steps to re-sort and/or re-calculate the results. One of the author's experiences as a decision analysis consultant led to the observation that in meetings the focus tended to shift to the "artifacts" of decision analysis (*e.g.* redrawing trees, redoing distributions), which was often disruptive to the rhythm of the free flow of ideas and the meeting process. This spreadsheet was developed to overcome this limitation, and thus developed "live" methods for producing tornado charts (which involved developing a "live sort" method), probability distribution sculpting, and decision tree creation and analysis. This example involves two files: da1.xls and da2.xls.

This fourth example, is based on the tool:

(10) Data tables

Method 10: Data Tables In our opinion, data tables are another useful but "non-intuitive" Excel feature. Data tables are a "what-if" tool that allows users to explore the effect of systematically changing values in a formula. Data tables are as close to a "do loop" as Excel gets (without using VBA). Data tables are best explained with an example (see table.xls and Figure 8).

The spreadsheet calculates a loan payment amount, for input variables such as interest rate, loan amount, etc. In the example, cells B3:B6 have input values, and cell B8 has the calculated loan amount. Cells A11:B20 are set up to be filled in as a one-variable data table. The column A12:A20 has possible interest rates. Cell B11 is a formula referring to the calculated payment amount. By selecting A11:B20, then going to **Data - Table** a menu appears. Since it is the interest rate that we want to vary, and that's in a column, enter B3 as the column input cell, then click "OK". The payment amount for each variable then fills the table. Note that the formula for a one-variable table goes at the top of the column to be filled in.

Cells D11:K20 are set up to be filled in as a two-way data table. The column D12:D20 again has possible interest rate values. The row E11:K11 has possible payment amounts.

FIGURE 8. Data Table Example

	A	B	C	D	E	F	G	H	I	J	K
1	Loan Payment Calculator										
2											
3	Annual Interest Rate:	8.00%									
4	# Payments per Year:	12									
5	Term of Loan (Years):	2									
6	Loan Amount:	$5,000									
7											
8	Payment:	$226.14									
9											
10	**One-Way Data Table**			**Two-Way Data Table**							
11	Interest Rate Effect:	$ 226.14		$226.14	$ 3,000	$ 3,500	$ 4,000	$ 4,500	$ 5,000	$ 5,500	$ 6,000
12	6.00%	$ 221.60		6.00%	$ 132.96	$ 155.12	$ 177.28	$ 199.44	$ 221.60	$ 243.76	$ 265.92
13	6.50%	$ 222.73		6.50%	$ 133.64	$ 155.91	$ 178.19	$ 200.46	$ 222.73	$ 245.00	$ 267.28
14	7.00%	$ 223.86		7.00%	$ 134.32	$ 156.70	$ 179.09	$ 201.48	$ 223.86	$ 246.25	$ 268.64
15	7.50%	$ 225.00		7.50%	$ 135.00	$ 157.50	$ 180.00	$ 202.50	$ 225.00	$ 247.50	$ 270.00
16	8.00%	$ 226.14		8.00%	$ 135.68	$ 158.30	$ 180.91	$ 203.52	$ 226.14	$ 248.75	$ 271.36
17	8.50%	$ 227.28		8.50%	$ 136.37	$ 159.09	$ 181.82	$ 204.55	$ 227.28	$ 250.01	$ 272.73
18	9.00%	$ 228.42		9.00%	$ 137.05	$ 159.90	$ 182.74	$ 205.58	$ 228.42	$ 251.27	$ 274.11
19	9.50%	$ 229.57		9.50%	$ 137.74	$ 160.70	$ 183.66	$ 206.62	$ 229.57	$ 252.53	$ 275.49
20	10.00%	$ 230.72		10.00%	$ 138.43	$ 161.51	$ 184.58	$ 207.65	$ 230.72	$ 253.80	$ 276.87
21											

Here the reference for the output cell goes in the "corner" of the table (cell D11 here). Select D11:K20, enter B3 as the column input cell and B6 as the row input cell, and click OK. A payment table is produced.

The "live decision analysis" spreadsheets involve many of the "top ten" methods in the paper. This model takes as given a fairly typical financial model to be used as the basis for deterministic evaluation of scenarios. It then uses data tables (which respond "live" to changes in their underlying formulas) and graphical displays to examine the effect of changes on the input data.

Tornado charts involving displaying the effects of changes in input values from the largest to the smallest impact (see Figure 9), and so "live" tornado charts require a "live" sorting procedure as well. The "live" sort relies heavily on rank and index functions (which are in the same family as the lookup functions previously discussed). The "live" probability distributions (see Figure 10) use mostly the same functions, and from them, we can get also calculate value of information in real-time. The "live" decision tree requires pivot tables as well. Once values for the endpoints of a decision tree are calculated, they are entered (not live) into a pivot table along with information about the sequence of events leading to each endpoint. Then, the process of "flipping the tree" — applying Bayes' rule to calculate conditional probability distributions under states of information — requires only the intuitive step of dragging columns so that they are in the same order as the event nodes in the version of the decision tree to be evaluated.

FIGURE 9. Live Decision Analysis Tornado Chart

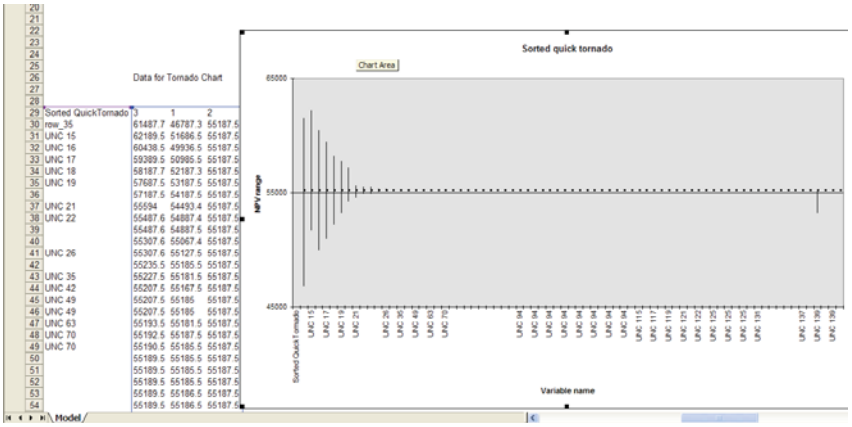


FIGURE 10. Live Decision Analysis Probability Distribution

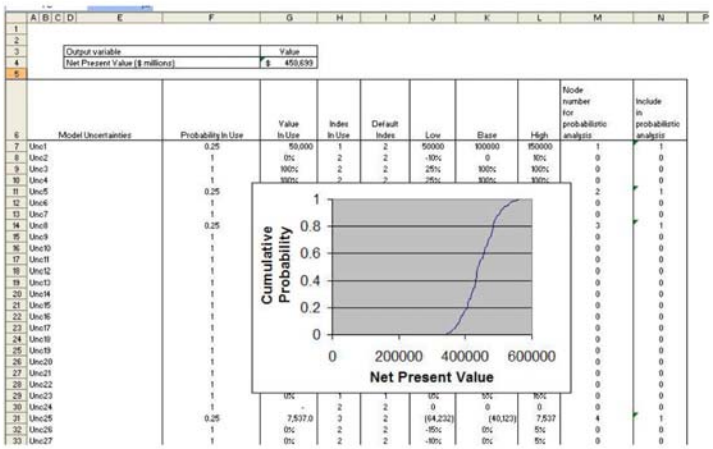


Figure 10: Live Decision Analysis Probability Distribution

Live decision analysis can change the focus from deterministic models — for which analysis is used to derive other values — to those derived values themselves (e.g., value of information, option value, and risk premium). By adjusting assumptions and decisions, it is then possible to actively sculpt a probability distribution. For example, a company might seek to maximize the value of information in a situation where it expects to have exclusive access to that information, or it might seek to maximize the risk premium in a situation where it has a higher risk tolerance than its competitors. This concept has facilitated rapid modeling for meta-decision making such as decision process design and risk allocation. The application described here is meant to support such efforts. It has been used in classroom settings, where students have found it to have intuitive appeal. As an aside, we undertook this and other efforts in part to apply spreadsheet techniques in our own field as a challenge in itself to learn more about the capabilities of Excel — in this case, to find a use for such capabilities as pivot tables and sort functions. Because Excel is a platform for application development, rather than merely an application itself, this kind of experimenting is an effective (and fun) way to develop skills.

7. Learn more! Join us! Help us "spread" the Good Word!

In this tutorial we have explored both "sin" and "salvation" in the spreadsheet kingdom. We have discussed ways to enhance the effectiveness and efficiency of the spreadsheet development process, including principles of spreadsheet engineering and robust spreadsheet design. We have discussed a number of good spreadsheet practices and discussed Excel features that support these practices. Highlighted among these practices in the examples are:

- the use of the methodology of: plan, build, test
- separation of data, from analysis, from presentation
- the creative mixing of multiple analysis methods and innovative presentation methods

The core of this tutorial however goes well beyond "tips and tricks" — the goal is to enable OR/MS professionals to harness the power of spreadsheets to support their particular areas of interest. Exploring spreadsheet functions and methods can spark new ideas for ways to

implement OR/MS methodology and systems, while in turn new OR/MS methods spark the need for more "killer ap" spreadsheets.

Spreadsheets are certainly not the only tool for OR/MS model development, and we would never advocate that all work be done in spreadsheets. However, the advantages of spreadsheets such as the ability to easily mix words, formulas, data, and graphs as well as their flexibility make them particularly appropriate for brainstorming and prototyping projects. One of the messages of this tutorial is, if spreadsheets are designed with purpose and care and if OR/MS developers take advantage of some of the advanced built-in (or added-in) capabilities, spreadsheets can be used for production applications as well.

If we have been successful with this tutorial, we have whetted your appetite for more. We encourage you to join SPRIG and become actively involved. Attend our sessions and conferences, share your own "killer aps", or even start your own spreadsheet research!

Acknowledgements The authors would like to thank Michael Johnson, whose editorial wisdom and keen eye have greatly improved this paper; the University at Albany and President Kermit Hall for their support of this endeavor; and SPRIG and Tom Grossman for focusing the attention of the OR/MS community on spreadsheets.

References

[1] S. C. Albright and W. L. Winston. *Spreadsheet Modeling and Applications: Essentials of Practical Management Science*. Southwestern College Publishing, Cincinnati, OH, 2004.

[2] A. A. Arens, R. J. Elder, and M. Beasley. *Auditing and Assurance Services: An Integrated Approach*. Prentice-Hall, 11th edition, 2005.

[3] B. Boehm and V. R. Basili. Software defect reduction top 10 list. *IEEE Computer*, 34(1):135–137, 2001.

[4] R. Butler and D. Chadwick. *Stop that Subversive Spreadsheet!* EuSPRIG, World Wide Web, `http://www.eusprig.org/eusprig.pdf`, 2003.

[5] R. J. Casimer. Real programmers don't use spreadsheets. *ACM SIGPLAN Notices*, 27(6):10–16, 1993.

[6] G. Croll. The Importance and Criticality of Spreadsheets in the City of London. In D. Ward, editor, *EuSPRIG 2005 Conference Proceedings*, pages 82–94, July 2005.

[7] Data Analysis Story Library (DASL). World Wide Web, `http://lib.stat.cmu.edu/DASL/Stories/dealersales.html`.

[8] EBI home page. World Wide Web, `http://www.webebi.com/`.

[9] EuSPRIG home page. World Wide Web, `http://eusprig.org`.

[10] T. A. Grossman, V. Mehrotra, and Ö. Ozluk. Spreadsheet information systems are essential to business. Working paper, School of Business and Management, University of San Francisco, San Francisco, CA, 2006.

[11] G. Harvey. *Excel 2003 for Dummies*. Wiley Publishing, Hoboken, NJ, revised edition, 2003.

[12] S. McConnell. *Rapid Development*. Microsoft Press, Redmond, WA, 1996.

[13] S. McConnell. *Code Complete*. Microsoft Press, Redmond, WA, 2nd edition, 2004.

[14] Microsoft Press Release, October 13, 2003. World Wide Web, `http://www.microsoft.com/presspass/press/2003/oct03/10-13vstoofficelaunchpr.mspx`.

[15] Microsoft Statistical Errors page. World Wide Web, `http://support.microsoft.com/default.aspx?kbid=828888\&product=xl2003`.

[16] P. O'Beirne. *Spreadsheet Check and Control*. Systems Publishing, Wexford, Ireland, 2005.

[17] S. G. Powell and K. R. Baker. *The Art of Modeling with Spreadsheets*. John Wiley & Sons, Danvers, MA, 2004.

[18] R. S. Pressman. *Software Engineering: A Practitioner's Approach*. McGraw-Hill, New York, NY, 6th edition, 2005.

[19] F. R. Jacobs R. B. Chase and N. J. Aquilano. *Operations Management for Competitive Advantage*. McGraw Hill/Irwin, New York, NY, 10th edition, 2004.

[20] Spreadsheet Research (SSR) Site. World Wide Web, `http://panko.cba.hawaii.edu/ssr/`.

[21] C. Ragsdale. *Spreadsheet Modeling and Decision Analysis*. Southwestern College Publishing, Cincinnati, OH, 4th edition, 2004.

[22] L. R. Rittenberg and B. J. Schwieger. *Auditing: Concepts for a Changing Enviroment*. South-Western College Publishing, 5th edition, 2004.

[23] C. Scaffidi, M. Shaw, and B. Myers. Estimating the Numbers of End Users and End User Programmers. In *IEEE Symp. on Visual Languages and Human-Centric Computing*, pages 207–214, September 2005.

[24] M. H. Serif, R. K. Ahuja, and W. L. Winston. *Developing Spreadsheet-Based Decision Support Systems Using VBA for Excel*. Duxbury Press, To appear 2006. see also: `http://www.innovativescheduling.com/publications/textbooks.asp`.

[25] I. Sommerville. *Software Engineering*. Addision-Wesley, Boston, MA, 7th edition, 2004.

[26] SPRIG home page. World Wide Web, `http://sprig.section.informs.org/`.

[27] J. Tennent and G. Friend. *Guide to Business Modelling*. Bloomberg Press, London, 2005.

[28] J. Walkenbach. *Excel 2003 Bible*. Wiley Publishing, Indianapolis, IN, 2003.