Building realistic forecasting models

Ray Hoare, Ph.D., HRS Ltd.

Introduction

I will be showing you a number of forecasting techniques that use existing information to predict future sales or costs. The examples will be calculated as I speak, using stand-alone programs that you can buy for a few thousand dollars.

Sometimes you will find similar methods implemented in large-scale systems designed to run many aspects of your business. These large-scale applications are much more expensive, and give you a lot of database-related tools for managing your business, with the smart application portion being relatively minor. I have had some strong negative comments from clients who have had to use some of these million-dollar applications for forecasting, but I have had no personal experience with them.

This talk has two sections: Time series analysis, and multi-variable modelling.

Time series analysis

In many cases you can predict the future sales by analysing past sales. Sometimes you just put a straight line through a plot of sales versus time. However, there are more sophisticated ways to do a similar thing, that use proven mathematical methods, and which will give you better answers.

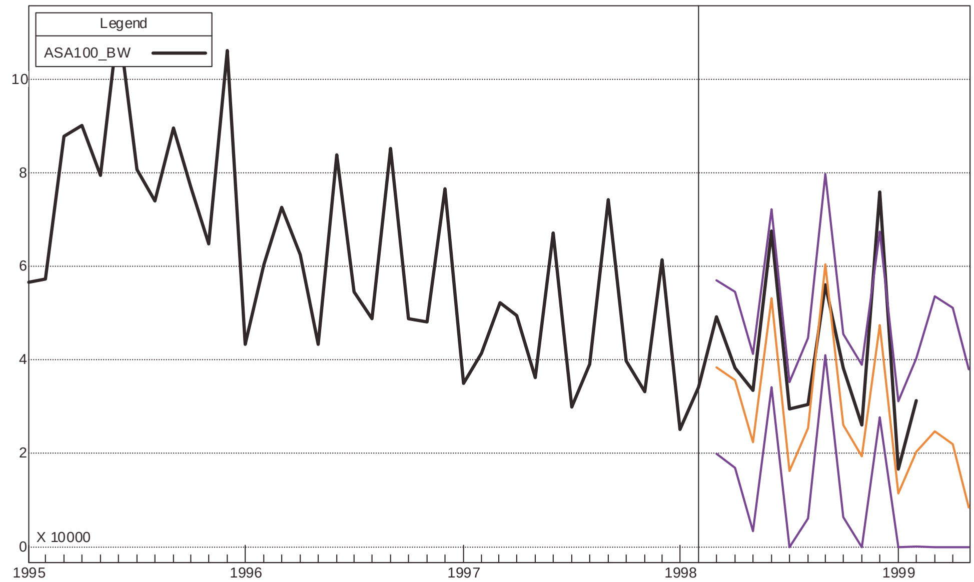
There are several components of the systematic variability of sales or other data with time:

- Level – the average value of the data
- Trend – does it go up or down consistently?
- Seasonality – is each month consistently high or low compared to the average?
- Events – are there predictable events in the historic data, such as sales or other marketing promotions
- Autocorrelation – can you be reasonably sure that one high (or low) value will be followed by another?

The level and trend are easy to work out with simple methods, but the seasonality, the autocorrelation and the effect of events need more work. There are a number of ways to do this, and some series are better described by different sorts of models than the ones that work well for other series.

Even once you pick the best kind of model, there are always different ways to choose the numbers to use in that model. Because of this complexity in the model design, it is almost always best to let an expert system based program work out the best model.

Also, it is not sufficient to fit the existing data well – you want a model that will predict well. Your software should be able to choose between a large number of methods – some programs only include a few of the possible methods and there is no "best" method.



In addition to those needs, it is common to have other complicating factors – you may be best able to forecast the total amount of a product, say beer volume, but need also to forecast the different categories, say to order packaging materials such as bottles and cans. The products you sell may have too short a life individually to work out seasonality, but you may be able to treat different models as a group to work out the necessary factors. (For example, printers or photocopiers have about an 18-month sales life.) Well-designed software makes dealing with these issues easy.

Example

The diagram shows one time series evaluated by Forecast Pro. The program has chosen a particular kind of ARIMA or Box-Jenkins model in this case. I have asked it to not use the last 12 months of data in the forecasting, so I can compare the forecast for that period with the actual results. I have calculated the forecast for 15 months. In the "forecast" period there are four lines – from the top down they are: the upper 95% confidence limit, the actual value, the most likely forecast, and the lower 95% limit.

The series has an obvious trend, and because the forecast goes up and down it must have seasonality. The fact that the program chose the Box-Jenkins method means there is a significant amount of autocorrelation, although this is always invisible to the eye.

Because Forecast Pro uses statistically sound methods, it can make realistic estimates of the confidence limits on the forecast. If you were using the data to estimate sales for the company budget you would use the most probable value. However, if you need to use it to order stock, you might use the upper 95% confidence limit – this would reduce the chance of running out of stock.

The kind of forecasting typified by the example is fine if you have only a few high-value items, or if you are applying the methods to the total sales of each division. If you are in a business with many relatively low-value products you may wish to use an approach in which thousands of forecasts are done individually and then combined – this is easily done where it is needed.

Uncertainty

In the modelling with Forecast Pro it was pointed out that the program predicted not only the most likely value, but it also predicted confidence limits – the range within which the forecast value will be 95% of the time. Whenever you get into forecasting you need to know the confidence of the forecast. This involves statistical concepts that some people find difficult, but it is essential to bring them into your decision-making process.

For instance, some modelling techniques were recently applied to the release of a film in NZ. The most likely outcome would have been a significant profit, but the modelling showed there was a 30% chance of a loss. The distributor chose not to go ahead, given that probability.

Multi-variable modelling

Multi-variable modelling tries to find relationships between the thing you need to know, and other data that is more easily controlled or predicted. You can search for statistical relationships, or you can build models from prior knowledge. Very often, the model is built in a spreadsheet.

Whenever you build a multi-variable model, you need to know how dependable it is. Simulation methods can allow you to do this.

Once you have a model, you have a method of saying "what if?". You can test sensitivity of your assumptions, or you can find the best value for the parts of the model that you can control.

Simulation with @RISK

Nett Present Value (10%) $363,263 The main output cell - summarises row 14

	Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
6	**Cash Flow**										
7	Total Revenue	$0	$0	$174,375	$145,078	$244,331	$321,435	$507,385	$711,839	$748,931	$787,877
8	Cost of Goods Sold	$0	$0	$69,750	$61,031	$102,533	$134,574	$211,954	$296,735	$311,572	$327,151
9	Gross Margin	$0	$0	$104,625	$84,047	$141,799	$186,861	$295,431	$415,103	$437,358	$460,726
10	Operating Expenses	$100,000	$130,000	$55,000	$30,000	$20,000	$20,000	$20,000	$25,000	$25,000	$25,000
11	Earnings Before Taxes	-$100,000	-$130,000	$49,625	$54,047	$121,799	$166,861	$275,431	$390,103	$412,358	$435,726
12	Tax Basis	-$100,000	-$230,000	-$180,375	-$126,328	-$4,529	$162,331	$275,431	$390,103	$412,358	$435,726
13	Income Tax	$0	$0	$0	$0	$0	$74,672	$126,698	$179,447	$189,685	$200,434
14	Net Income	-$100,000	-$130,000	$49,625	$54,047	$121,799	$92,188	$148,733	$210,656	$222,674	$235,292
16	**Market Conditions**										
17	Number of Competitors			0	1	1	1	1	1	1	1
18	Unit Cost			$23.25	$24.41	$25.63	$26.91	$28.26	$29.67	$31.16	$32.72
19	Inflation Rate			5%	5%	5%	5%	5%	5%	5%	5%
20	Tax Rate	46%	46%	46%	46%	46%	46%	46%	46%	46%	46%
22	**Sales Activity**										
23	Sales Price			$58.13	$58.03	$61.08	$64.29	$67.65	$71.18	$74.89	$78.79
24	Sales Volume			3000	2500	4000	5000	7500	10000	10000	10000
26	**Production Expense**										
27	Product Development	$50,000	[illegible]	[illegible]	[illegible]	$ -	$ -	$ -	$ -	$ -	$ -
28	Capital Expenses	$50,000	$100,000	$25,000	$10,000	$ -	$ -	$ -	$ -	$ -	$ -
29	Overhead	$0	$10,000	$20,000	$20,000	$20,000	$20,000	$20,000	$25,000	$25,000	$25,000
30	Total Expenses	$100,000	$130,000	$55,000	$30,000	$20,000	$20,000	$20,000	$25,000	$25,000	$25,000

=RiskNormal(50000,10000)

Calculating the profitability of an investment

We will illustrate modelling in a forecast context by taking a fictitious company that wishes to build widgets. It has to spend money developing the design for a start, and then when it starts to sell them it has to work out how many it can sell and what effect competition has on the sales volume and price.

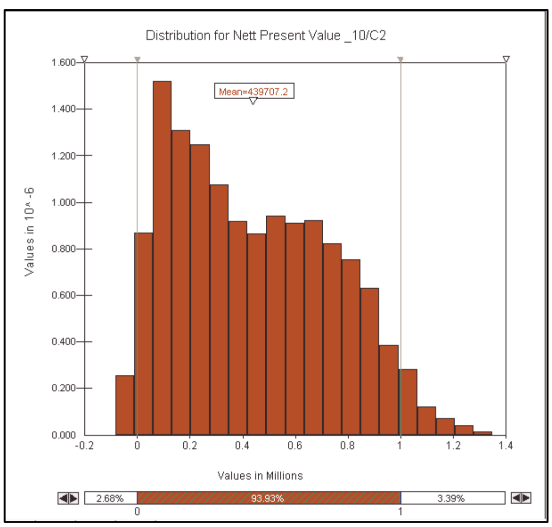
The cost of development is a big unknown, of course, and the @RISK modelling add-in for Excel allows you to specify a distribution of values. Perhaps it is randomly but “normally” distributed around a most likely value. Perhaps it is more likely to be above the most probable value than below it. Whatever you decide, you can set a spreadsheet cell to take the appropriate range of values.

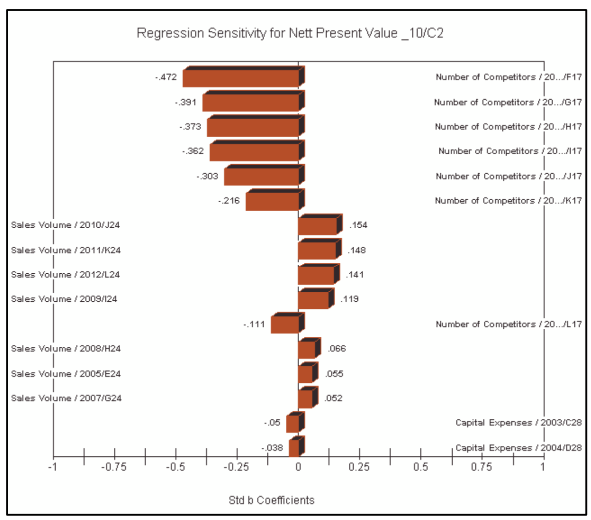
In the model that will be demonstrated, (shown above) there is an equal likelihood of competitors arriving or leaving each year.

The output of the model is related to those cells that calculate the total value of the investment – year-by-year, or overall for a 10-year period.

Once the model is set up, it is recalculated many times. Each re-calculation cycle the program chooses different values for the cells containing distribution functions, and the resulting values of the output cells are stored. You can then look at the values individually, or graph them.

The picture above shows a spreadsheet that has @RISK functions in the shaded cells. The comment bubble shows the function used to describe the product development cost. Other coloured cells have different expressions for the variable values.





When the model was recalculated the values in the NPV cell were recorded, and the distribution of those values is shown to the left, above. There is a 2.7% chance of losing money, and a 3.4% chance of making more than a million dollars.

Sensitivity of the model

When you run models like this you never stop at the first attempt. @RISK provides tools that let you test which assumptions of the model are most important – in this case it is the number of competitors. (See the tornado plot at the right.) It works out that with this assessment of the competition you don't actually have to worry much about the development costs!

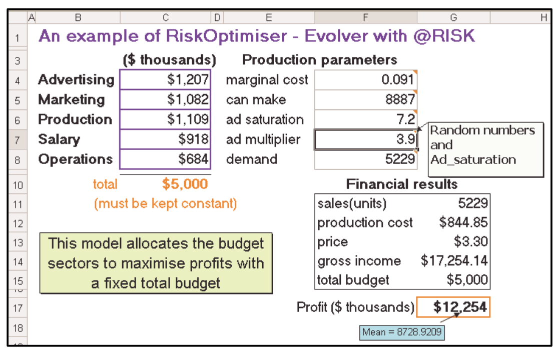
Conclusion on risk modelling

If you have forecasting tasks that involve building a model, you will always have uncertainty. Monte-Carlo analysis with @RISK or similar products will allow you to make realistic assessments of the uncertainty in your forecasts, so you can decide whether to go ahead right away, or whether you need more data on critical aspects of your model.

Optimising

Optimising is a stock-in-trade of the Operations Researcher. You often have a choice of things you can do, such as where to put resources. Optimising allows you to choose a mix that will give you the best profit. There is a tool for doing this in modern spreadsheets, but it is very limited in the range of problems it can tackle, and is very slow compared to some methods.

We will run an optimisation model that allocates resources among departments in a company, while still allowing for uncertainty.



The model, shown to the right, is based on relationships between expenditure on marketing and advertising and the sales of the company. These of course involve uncertain values.

The program makes a guess at a good combination of expenditure in the 5 departments, and works out a Monte-Carlo simulation to find the most likely profit. It then takes another guess, and if it is a higher profit uses the values in this better guess as a basis for a yet better guess. The method is known as Genetic Algorithm optimisation.

A panel shows the progress of the simulation, and the properties of the different solution sets.

Summary

I have shown just a few of the techniques that can be used to extract information from existing data, to come up with estimates of the most likely outcomes of business scenarios, and to estimate the range of variability around those most likely outcomes.

Although the forecasts are based on sophisticated mathematical methods, with the aid of a computer you can apply these methods without a detailed knowledge of the mathematics, nor any awareness of the restrictions and pitfalls that await the unwary.

Unless you are a specialist analyst it is essential to have access to someone who can keep an eye on what you are doing, and warn you of things you should not do, or better ways to implement the methods. You may hire them if you are in a large company, but you can find these people - who specialise in “operations research” - in universities, some of the better consulting companies, and possibly in the software vending companies. Use them!